

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO,

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NCL; NEU; HER-2
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <BE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: intracellular #status predicted <INT>
 F;726-734/Region: protein kinase homology <KIN>
 F;68,124,187,259,530,571,623/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.58; Score 6672; DB 1; Length 1255;
 Best Local Similarity 98.24; Pred. No. 2.2e-265;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASQTQCT-----QYIKANSKFITIGITELLYQGCVVQGNL 54
 DB 1 MELAALCRWGLLLALLPPGAASQTQCTGCTDMKRLRLPASPTHLDMRLHLHYQGCVVQGNL 60

QY 55 ELTYLPTNASLFLQDIQEQGVYLIAHNQVRQVQLRILVRGTQLFEDNYALAVLDNG 114
 DB 61 ELTYLPTNASLFLQDIQEQGVYLIAHNQVRQVQLRILVRGTQLFEDNYALAVLDNG 120

QY 115 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 174
 DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 180

QY 175 LTLIDNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGKPLPTDCCHEQC 234
 DB 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGKPLPTDCCHEQC 240

QY 235 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLVCPHMQEVAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 354
 DB 301 YNYLSTDVGSCTLVCPHMQEVAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360

QY 355 IOEFAGCKKIFGSLAFLPSFDGDPASNTAPLOEQLOVFETLEETIGVLYISAWPDSL 414
 DB 361 IOEFAGCKKIFGSLAFLPSFDGDPASNTAPLOEQLOVFETLEETIGVLYISAWPDSL 420

QY 415 DLVSFQNLQVIRGILHNGAYSITLQGLGISWGLSLRELGSGLAIHHNTHLCFVHTV 474
 DB 421 DLVSFQNLQVIRGILHNGAYSITLQGLGISWGLSLRELGSGLAIHHNTHLCFVHTV 480

QY 475 PMDQLFRNHQALHTANRDECEVGLACHOLCARGHGWGPGTQCNCQSOFLRGQC 534
 DB 481 PMDQLFRNHQALHTANRDECEVGLACHOLCARGHGWGPGTQCNCQSOFLRGQC 540

QY 535 VEECRVLQGLPREYVNRHCLCPHPECQPCQNGSVTCFGEADOCVCAHYKDPFFCVARC 594
 DB 541 VEECRVLQGLPREYVNRHCLCPHPECQPCQNGSVTCFGEADOCVCAHYKDPFFCVARC 600

QY 595 PGVVKPDLSPYMPKPPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSISAVVG 654

RESULT 2

148161
 p-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Y.;
 Gene 140 251-255, 1994
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48161; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:g493236; PIDN:BA03801.1; PID:g747595
 C;Genetics:
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.5%; Score 5861.5; DB 2; Length 1254;
 Best Local Similarity 86.1%; Pred. No. 2.6e-232;
 Matches 1081; Conservative 59; Mismatches 108; Indels 7; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASQTQCT-----QYIKANSKFITIGITELLYQGCVVQGNL 54
 DB 1 MELAALCRWGLLLALLSPGASQTQCTGCTDMKRLRLPASPTHLDIRHLHYQGCVVQGNL 60

QY 55 ELTYLPTNASLFLQDIOEVQGVVLIHAHQVROVPLQRLIRVIRGTOLFEDNYALAVLDNG 114
DB 61 ELTYLPANATLSFLQDIOEVQGVVLIHAHQVROVPLQRLIRVIRGTOLFEDNYALAVLDNR 120
QY 115 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPCOLCYQDTILWKDIFHKNNOLA 174
DB 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIIRGNPCOLCYQDTILWKDIFRKNQOLA 180
QY 175 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQC 234
DB 181 PVDIDTNRSRACPPCAPACKDNHCWGPEDCQTLGCTIAPRAVPAARARLPTDCCHEQC 240
QY 235 AAGCTGPKGSDCLACLFHNSHSGICELHCPALVYNTDTFESMNPGRYTFGASCVTAC 294
DB 241 AAGCTGPKGSDCLACLFHNSHSGICELHCPALVYNTDTFESMNPGRYTFGASCVTTC 300
QY 295 YNYLSTDVGSCTILVCPILHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLREVAVTSA 354
DB 301 YNYLSTEVGSCTILVCPILHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLRGARAITSA 360
QY 355 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFTLEETIGLYISAWPDSLP 414
DB 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFTLEETIGLYISAWPDSLSLH 420
QY 415 DLSVFQNLQVIRGRIILHNGAYSITLQGLGISWGLSLRSLGSLGLIHHNTHLCFVHTV 474
DB 421 DLSVFQNLVIRGRVLHDGAYSIALQGLGIRWLGSLRSLGSLGLIHRNTHLCFVHTV 480
QY 475 PWDLFRNHQALHTANRDESCVGEGLACHOLCARGHCWGPQTQCVNCSOFLRGQC 534
DB 481 PWDLFRNHQALHTANRDESCVGEGLACHOLCARGHCWGPQTQCVNCSHFLRGQC 540
QY 535 VBCRYLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVAC 594
DB 541 VKECRWKGLPREYVNGKCLPCHPECPQNSVTCFGEADQCTACPHYKDSPPFCVAC 600
QY 595 PSQVKPDLVMPYKWPFDDEGACQPCPNCTHSCVDLDKGCBAEQASPLTSIVSAVVG 654
DB 601 PSQVKPDLVMPYKWPFDDEGACQPCPNCTHSCVDLDKGCBAEQASPLTSIVATVVG 660
QY 655 ILLAVVLGVFGILLIKRQOKIRKYTMRLLQTELVEPLTPSGAMPNQAQRIKTEL 714
DB 661 ILLFLVGVVGLIKRQOKIRKYTMRLLQTELVEPLTPSGAMPNQAQRIKTEL 720
QY 715 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 774
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGLGP 780
QY 775 YVSRLLGICLTSTVQLVTPQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGMVLYEDVR 834
DB 781 YVSRLLGICLTSTVQLVTPQMLPYGCLLDHVRENRGLSGODLLNWCQIAKGMVLYEDVR 840
QY 835 LVHRDLAARNVLKSPNHVKITDPLGLARLLIDETEHADGGKVPKWALESILRRRT 894
DB 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLIDETEHADGGKVPKWALESILRRRT 900
QY 895 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVYKWM 954
DB 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVYKWM 960
QY 955 IDSECRPRFELVSEFSRMARDQRFVVIQNEDLGPASPLDSTFYRSLDEDDDMGLVDA 1014
DB 961 IDSECRPRFELVSEFSRMARDQRFVVIQNEDLGPASPLDSTFYRSLDEDDDMGLVDA 1020
QY 1015 EYLVPOQGFCCDDPAPGAGWVHHRSSSTSGGDLTLGLEPEEERAPSLAPSEG 1074
DB 1021 EYLVPOQGFCCDDPAPGAGWVHHRSSSTSGGDLTLGLEPEEERAPSLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPTHDPSPLQRYSEDPTVPLPSSTDCGVAPLTCSPQPEV 1134
DB 1081 AGSDVFEGLGMAATGKQPSISPRDLSPLQRYSEDPTVPLPSSTDCGVAPLTCSPQPEV 1140
QY 1135 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNVGVKDVFAFGGAVENPEYLTQ 1194

DB 1141 NOPEVRPQPLTPEGLPPVRPAGATLERPKTSLSPGKNVGVKDVFTFGGAVENPEYLVPR 1200
QY 1195 GGAAPQPHPPAFSPAFDNLYYWDODPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1249
DB 1201 GGSASQPH-PPALCPAFDNLYYWDODPSRGSPPNFTFGTPTAENPEYLGLDVVP 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (EC 2.7.1.112) new precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The new oncogene encodes an epidermal growth factor receptor-related protein.
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, Carcinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no 2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <MAS>
A>Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho-
F:1-13/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 86.4%; Score 5854; DB 1; Length 1260;
Best Local Similarity 86.1%; Pred. No. 5.4e-232;
Matches 1082; Conservative 54; Mismatches 113; Indels 8; Gaps 3;
QY 1 MELALCRWGLLALLPQAASTQVCT-----QVIAKSKFIFGITELLYQGCQVVOGNL 54
DB 4 MELANCRWGLLALLPQIACTQCTGDMKRLRPLASPETHLDMRLHYQGCQVVOGNL 63
QY 55 ELTYLPTNASLFLQDIOEVQGVVLIHAHQVROVPLQRLIRVIRGTOLFEDNYALAVLDNG 114
DB 61 ELTYLPANATLSFLQDIOEVQGVVLIHAHQVROVPLQRLIRVIRGTOLFEDNYALAVLDNR 120
QY 115 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPCOLCYQDTILWKDIFHKNNOL 173
DB 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIIRGNPCOLCYQDTILWKDIFRKNOL 183
QY 175 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCCAGCARCKGPLPTDCCHEQ 233
DB 181 PVDIDTNRSRACPPCAPACKDNHCWGPEDCQTLGCTIAPRAVPAARARLPTDCCHEQ 243
QY 235 AAGCTGPKGSDCLACLFHNSHSGICELHCPALVYNTDTFESMNPGRYTFGASCVTAC 293
DB 241 AAGCTGPKGSDCLACLFHNSHSGICELHCPALVYNTDTFESMNPGRYTFGASCVTTC 303
QY 295 YNYLSTDVGSCTILVCPILHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLREVAVTSA 353
DB 301 YNYLSTEVGSCTILVCPILHNOEVTAEDGTQCEKSKPCARVCYGLGMEHLRGARAITSD 363

A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
Cell 59, 33-43, 1989
A;Title: Functional independence of the epidermal growth factor receptor from a domain I
A;Reference number: A33331; PMID:900032333; PMID:2790960
A;Contents: Annotation: internalization signal
C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
C;Genetics:
A;Gene: GDB:EGFR
A;Cross-references: GDB:l20610; OMIM:l31550
A;Map position: 7p12.3-7p12.1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-1210/Product: EGF receptor #status predicted <MAT>
F;25-645/Domain: extracellular #status predicted <EXT>
F;75-300/Domain: EGF receptor extracellular domain repeat <EE1>
F;390-600/Domain: EGF receptor extracellular domain repeat <EE2>
F;646-668/Domain: transmembrane #status predicted <TM>
F;669-1210/Domain: intracellular #status predicted <INT>
F;710-975/Domain: protein kinase homology <KIN>
F;718-726/Region: protein kinase ATP-binding motif
F;999-1046/Region: coated-pit mediated internalization signal
F;1047-1210/Region: inhibitory
F;128.175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
F;745/Active site: Lys #status experimental

Query Match	46.3%	Score 3135;	DB 1;	Length 1210;
Best Local Similarity	49.4%	Pred. No. 5.3e-121;		
Matches	626;	Conservative 179;	Mismatches 327;	Indels 114; Gaps 23;
QY	11	LLAALLPPGAA--STQVC-----TQYIKANSKFIGITELLYOGCVOVGNLELTLYLPT	61	
Db	14	LLAALCPASRALBEKKVCOGTSNKLTLQGTFFDHFLSL-QRMFNNECVVLGNLEIITVVQR	72	
QY	62	NASLSFLQDIQEQVQGYVLIAHNOVRQVPLQRLRIVRGTQLPFEDNYALAVLDNGDPLNNTT	121	
Db	73	NYDSLFLKTIQEVAGYVLIALTVERIPLNLIIRRMVYENSVALAVLSNYD-----126		
QY	122	PVTGASPGGIREIQLRSLTILKGGVLIQRNPQLCYODTILWKDI FHNQOLAULTLIDTN	181	
Db	127	----ANKTGELKEIPLMRNLQBIHLGAVRFSNNPALCNVIESIQWRDIVSDFLSNNSMD	182	
QY	182	RSRACHPCSPWKGSRGWSSEDCQSLTRTVCAGGCA-RCKGPLPTDCCHEQCAAGCTG	240	
Db	183	HLGSCQKDCSPNGSCWGAAGEENCQKLKILIIAQQCSGRGRKSPSDCCCHNQCAAGCTG	242	
QY	241	PKHSDDCLACILFHNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTACPYNLST	300	
Db	243	PRESDCIVCKFRDEATCKDTCPLMLYNPTYQMDVNPESKYSFGATCKVKCPRYNVT	302	
QY	301	DVGSCTILVCPHINQEVTAEDGTORCEKSPCARVCYGLGMEHUREVRVATSIANIQEFAG	360	
Db	303	DHGSVCRAACGADSYEM--EEDGVKRCCKEGPCRKVCNGIGIGEFKDSLISNATNIRKFKN	361	
QY	361	CKKIFGSLAFLPSFDGDPASNTAPLOEOLQVFTLEEITGYLIISAWPDSLPLDSVFQ	420	
Db	362	CTSIISGDHLIIPVAFRGDSFTHPPDQBELDILKTVKEITGFLIIQAWPENRTDLHAFE	421	
QY	421	NLOVIRGRILHNAGYSITLQGLGISWLGRLSRELGSGLALIIHNHTLCHFVHTVPWQLF	480	
Db	422	NLEIIRGRTKQHGQFSLAVVSLNITSLGLRSLKEISDGDVTISGNKNLCYANTINWKKLF	481	
QY	481	RNPQOALLHTANRPEDECVCEGLACHOLCARGHCWGPGPTQCVNCQFRLQOECVEECRV	540	
Db	482	GTSGQTKTIIISNRGENSKATQGVCHALCSPGECWGPPEPRDCVSRNVSRGRCVDRCKL	541	
QY	541	LOGLPREYVNAHCLCPHCEPCQONGSVTQFGPEADQCVAHAHYKDPFPFCVARCPSGVKP	600	
Db	542	LEGEPRFVENSECIIQCHPECLFOAMNITCTGRGPDNCIQCAHYIDGPHCVKTCFAGVMG	601	
QY	601	DLSYMTIWKPPDEBEGACQPCPNCTHSCVDLDDKGCPAERQASPLTISVSAWG---ILL	657	

RESULT. T 5

epidermal growth factor receptor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 06-Jan-1995 #sequence.revision 06-Jan-1995 #text change 18-Jun-1999
C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
R;Juetterke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins
Genes Dev. 8, 399-413, 1994
A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF r
A;Reference number: A53183; MUID:94170986; PMID:8125255
A;Accession: A53183
A;Molecule type: mRNA
A;Residues: 1-1210 <LUE>
A;Cross-references: GB:U03425
R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A;Title: Comparison of EGF receptor sequences as a guide to study the ligand bind
A;Reference number: A43818; MUID:91232866; PMID:2030916
A;Accession: A43818
A;Molecule type: mRNA
A;Residues: 1-714 <AVI>
A;Cross-references: GB:X59698
R;Eisinger, D.P.; Serrero, G.
submitted to the EMBL data Library, June 1992
A;Reference number: S24942
A;Accession: S24942

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
A:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-1223/Product: epidermal growth factor receptor #status predicted <SIG>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.5%; Score 3085.5; DB 1; Length 1223;
Best Local Similarity 48.4%; Pred. No. 5.6e-119;
Matches 628; Conservative 175; Mismatches 342; Indels 153; Gaps 27;
QY 8 RWGLLLALLPPGAA-----STQVC-----TOYIKANSKFIGITELYQCVQVGNL 54
DB 13 RGAALVLLLLGVALCSAVEEKVCGQTNNKLTQLGHVEDHFTSL-QRMNNCEVLSNL 71
QY 55 ELTYLPTNASLSFLQDIOBQGVGLIAHNVQVQLRIRIVRGTLQFEDNVALAVLDNG 114
DB 72 EITYVEHNDLFLTKTQEVAGVGLIALNMVDIPLNIQIIRGNVLYDNSFALAVLSNY 131
QY 115 DPLNNTPTVTGASPGRLQLRLSLTEILKGGVLIORNPOLCVQDTILWKDIFHKNNOLA 174
DB 132 H-MNKTQ-----GLRELPMKLSLEILNGGVKLSNPKLNCMTVLNLDIIDSRL-P 181
QY 175 LTLID-TNRSRACHPCSPCKSGRWBSSEDCQSLTRTVACGGA-RCKGPLPTDCCHE 232
DB 182 LTVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCKGVPSDCCHN 241
QY 233 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTA 292
DB 242 QCAAGCTGPRSDCLACRFRDADTKCTCPPLVLYNPTYQMDVNPCKYSGATCVRE 301
QY 293 CPYNYLSTDVSGCTLVCPLNHEVTAEDGTQCEKSKPCARVCYGLGMHLREYRAVTS 352
DB 302 CPNYVVTDHGVSRCNSNTDTVEV-ENGVRKCKKCDGLCKVCVNGIGIGELKGLISNA 360
QY 353 ANIQEFAGCKITGSLAFIPESFGDGPASNTAPLOEQLOVFTLEITGYLYISAWPDS 412
DB 361 TNIDSPKNTKINGDVSILPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQAWPDN 420
QY 413 LPDLSVFQNLQVIRGILHNGAYSILTQGLGTSWGLSLRLGSLGLALIHNTLCFVH 472
DB 421 ATDLYAFENLEITIRGTRKQHGQYSLAVNLKIQSLGLSLKESISDGTALMKNKRLCYAD 480
QY 473 TVPWQDLFRNPHQALLHTANRDECEVGBGLACHOLCARGHCWGPQTCVNCOSFLAQO 532
DB 481 TNWNSLFATQSQKTKIQNRKNKNDCTADRHVDCPLSDVGCWGPQGFHCFSRFSRQK 540
QY 533 ECVEECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGPEADQCACAHYKDPFF 589

DB 541 ECVKQCNIQLQGEPRFERDSEKCLPCHSECLVQNSTAYNTTSCGPGDHCMKAHFIDGPH 600
QY 590 CVARPCSPGVKPDLSVMPYWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIV 649
DB 601 CVKACPAVGLGENDTL-VWKYADANAVCOLCHPNCTRGCKGPGLEGCP---NGSKTPSTA 656
QY 650 SAVV-GILLVVVLGVVFGILKRRQOKIRKYTRMLLQSTELVELFPLTSGAMPNOQMRI 708
DB 657 AGVWGGLLCLVVVGLGIGLYLRRR-HIVKRTLRLLQERELVEPLTPSGEAPNQAHLAI 715
QY 709 LKETELRKVKVLGSGAFGTIVYKGIWIPGENVKIPVAIKVLRENTSPKANKILDEAYYM 768
DB 716 LKETEFKVKVLGSGAFGTIVYKGIWIPGENVKIPVAIKVLRENTSPKANKILDEAYYM 775
QY 769 AGVGPYVYSRLIGLICLTSTVOLTPOLMPYVCLLDHVRNRRGLSGQDLNLCWQIAKMS 828
DB 776 ASVDNPHVCRLLIGLICLTSTVOLTPOLMPYVCLLDHVRNRRGLSGQDLNLCWQIAKMS 835
QY 829 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESI 888
DB 836 YLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGGKVPKMALESI 895
QY 889 LRRRTHOSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPPICITDVYMI 948
DB 896 LHRIYTHOSDVWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPOPPICITDVYMI 955
QY 949 MYKCMWIDSECRPRELVSFRSMARDPQRFVLIQ-NEDLGPASPLDSTFVRSLLDDEDD 1007
DB 956 MYKCMWIDSECRPRELVSFRSMARDPQRFVLIQ-NEDLGPASPLDSTFVRSLLDDEDD 1015
QY 1008 MGDVDADEYLVPQOGFFCPDAPAGAGGVHHRSSSTRSGGDLTLGLLEPSEBEAPRS 1067
DB 1016 MEDIVDADEYLVPQOGFF-----NSPST-----SRT 1041
QY 1068 PL-----APSEGAGSDVFDGLGMGAAGKLSLPHDPSLPQRYSEDPTVPLPSET--DG 1120
DB 1042 PLLSLSLSATSNNSATNCID-----RNGQCHPVREDSFVQRYSSDPTGNFLESSIDDG 1093
QY 1121 VYAPLTCSPQEVYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVP- 1179
DB 1094 FL-----PAPEYVQ--LMPKKPS-----TAMVOQIYNNISL 1124
QY 1180 -----AFGGAVENPEYLTQGGAAAPQHPHPPAFSPAFDNLVYWDQ----- 1219
DB 1125 TAISKLPMDSRVQNSHSTAVDNPEYL-----NTNOSPLAKTVFESSPYIQSGNHQI 1176
QY 1220 --DPPE-----FGAPPSTFKGTPTAENPEYLGLDVP 1248
DB 1177 NLDNDPYQQDFLFPNETKPNGLLKVPAAENPEYLRVAAP 1214

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Flowerman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Nee,
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

```

Query March          43.8%, Score 2968.5; DB 2; Length 1308;
Best Local Similarity 45.1%; Pred. No. 3.6e-114;
Matches 609; Conservative 186; Mismatches 369; Indels 185; Gaps 30;

Qy      9  WGLLALLPPGAAA---STOVCTQYIKANSKFIGITEL-----LYQCQCVVQCNLE 55
Db      8  WVVSLIVAAGTVQPSQSVC---GTENKLSLSDLSEQOYRALRKYKYEVCVVMGNLE 64

Qy     56  LTYLPTNASLFLQDIQEVQGYVLIHAHQVRQVQLRLRIVRGTQQLPEDNYALAVLNDGD 115
Db     65  ITSIEHRNRLSFLASREVETGYVLVALNQFRYLPLENLRIIRGTIKLYEDRYALAIPLNYR 124

Qy    116  PLNNTTPTVGTASPGGLRELRQLRSUTEILKGVLLQRPOLQCYQDTIILWKQIFHKNQIAL 175
Db    125  KDGNF-----GLOELGGLKNLITELINGGVYVDONKFLCYADTIHWQDIIVRNPWPSNL 175

Qy    176  TLIDNRSRACHPCSPMKCSRWGSESDCOSLTRTVCAAGC--ARCKGGLPTDCCHEQC 234
Db    176  TLVSTNGSGGGRCHKSTG--RCWGPTEHHCQTUTRVCAQCQDGRCYGPVYSDCCHREC 234

Qy    235  AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFBSMPNPEGRYTFGASCVTACP 294
Db    235  AGGCGPKDTCDFACMFNDSGACVTCQPQTFFVNPFTTFOLEHNFNAKYTYGAFCVKCKP 294

Qy    295  YNYLSTDVGSCTLVCPLIHQNVETADGTQRCCKSPCARVCYGLGMEHLREVRVNTSAN 354
Db    295  HNFV--VDSSSCVRAACPSSKMEV--EENGIKMKCPKCTDICPKACDGGTGTGLMSAQTVDSN 352

Qy    355  IQEPAQCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFTLEBEITGYLYISAWPSLIP 414
Db    353  IDKEINCTKINGNLIFIVTGIHGDPYNAIEAIDPEKLNVRTVREITGFLNIOSWPPNMT 412

Qy    415  DLSVFQNLQVIRGRILHNGAYSFLTQGLGISWLGSLRLSGLSGLALIHENTHLCFFVHT 474
Db    413  DFSVFSNLVTIGRVLVSGILSLILKQGITSLQFQSLKEISAGNIYITNSNLCCYHTTI 472

Qy    475  PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 534
Db    473  NWTTLFTSTINQRIVIRNRAENACTABGMVGNHLCCSSDGCWGPGDQCLSCRRFSRGRIC 532

Qy    535  VEECRVQLGLPREYVNAHCLPCHPEQCP--QNGSVTCFGEADOCVACAHYKDPFPCVAR 593
Db    533  IESCNLYDGBREFEFGSGICEVEDPQCEKMEDEGLLTCHGPGPDNCTKCSHFQDGPNCVEK 592

Qy    594  CPSGVKPDLSYMPYTWKFPDEBEGACQPCPINCTHSCVDLDDKGC-----PASQRA 642
Db    593  CPDGLQANSF--IFKYADPDRECHPCPNCTQCGNGPTSHDCIYYPTWGHSTLPQHAR-- 649

Qy    643  SPLTISVSAVV--GILLVVVLGWVFGILIKRRQQKIRKYTMRRLLQETELVEPLTPSGAMP 701
Db    650  TPL--IAAGVIGLFLIVLGLTFEAVVRRKSIK--KKRALRRFL--ETELVEPLTPSGTAP 705

Qy    702  NQAQMRILKETELRKVKVLGSGAGFTYVKGIIWPDGENVKIPVAIKVLRENTSPKANKEI 761
Db    706  NQAQLRIKETELKRVKVLGSGAGFTYVKGIIWVPEGETVKIPVAIKTLNETTGPKANVEF 765

Qy    762  LDEAVVWAGVSPVSVRLIGICLTSTVQLVTQMLPYCGLADHDVRENGRLSGODLLNWC 821
Db    766  MDEALINWASHMDHPLVRLGLVCUSPTQLVTQLMPHGCLLEYVHEHKDNIGSQLLNWC 825

Qy    822  QIAKMSYLEDLVLRHRLDAARNVLVKSPNHVKITDFGLARLIDIDETEYHADGGKVPK 881
Db    826  QIAKGMVYLEERLVRHRLDAARNVLVKSPNHVKITDFGLARLLEGEDEKEYNADGGKVPK 885

Qy    882  WMALESILRRRPTHQSDVSWGYVTWBLMTFGAKPYDGIIPAREIPDILLEYGERLPQPPIC 941
Db    886  WMALECITHYKFTHQSDVSWGYVTWBLMTFGGKPYDGIPTREIPDILLEYGERLPQPPIC 945

Qy    942  TIDVYIMVWCMMIDSCRPRFRELVEFSRMARDPQRFVVIQNE--LGPASPLDSTFYR 1000
Db    946  TIDVYIMVWCMMIDASRPFKELAAEFGRMARDPQRYLIVQGDHMKMLPSPNDSKFFQ 1005

Qy   1001  SLLEDDDMGDLVDAEEYLVPOQGFCCPDPAFGAGGMVHHRHSRSTSGGCDLTLGLEPS 1060

```

Db	1006	NLLDEELEDMDABEYLVP-QAFNIPPP-----IYTSRARIDSNRS-----EIGHSP 1053
Qy	1061	EEAPRS-----PLAP-SEGAGSDVFQDGLMGAAKG 1091
Db	1054	PAYTPMSGNFVYRDGGFAAEQGVSVPYRAPTSTIPAPVAQATATEIFDDSCNGTLRK 1113
Qy	1092	LQSLPTHDPSPLOYSDDPTVPLFS-----ETDGYVAPLTCSPOEYVQNPQVRP 1144
Db	1114	VPAPHVQEDSSQYSADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPE----- 1167
Qy	1145	SPREGPLPAARPAGATLERAKTILSPKNGVYKDVAFPGGAVENPEYLTPOGGAAPQHP 1204
Db	1168	---ENPFVSR---KNGDLQ-----ALDNPETHNASNG-----PP 1196
Qy	1205	PA-----FSPAFDNLVYWDQDPBERGA--PP 1228
Db	1197	KADEYVNEPLYNTFANTLGKAYLYKNLILSMPEKAKAFDNDPYWNHSLPPRSTLQHP 1256
Qy	1229	STFKGTPT-----AENPEYL 1243
Db	1257	DYLOEYSTKYFYKONGRIRPIVAENPEYL 1285
RESULT 8		
S06142		
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish		
N;Alternate names: epidermal growth factor receptor homolog; kinase-related tra		
C;Species: Xiphophorus maculatus (southern platyfish)		
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000		
C;Accession: S06142; S13809		
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeucler, W.; Raulf, F.; Telling, A.		
Nature 341, 415-421, 1989		
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-induci		
A;Reference number: S06142; MUID:90015140; PMID:2797166		
A;Accession: S06142		
A;Molecule type: DNA		
A;Residues: 1-1166 <WIT>		
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291		
R;Adam, D.; Maeucler, W.; Scharlt, M.		
Oncogene 6, 73-80, 1991		
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in X		
A;Reference number: S13807; MUID:91125882; PMID:1846957		
A;Accession: S13809		
A;Status: preliminary; translation not shown		
A;Molecule type: DNA		
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>		
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285		
C;Genetics:		
A;Gene: mrk		
A;Map position: Y		
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1		
C;Superfamily: epidermal growth factor receptor; protein kinase homology		
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane prot		
F;1-25/Domain: signal sequence #status predicted <SIG>		
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <		
F;707-972/Domain: protein kinase homology <KIN>		
F;715-723/Region: protein kinase ATP-binding motif		
Query Match 39.6%; Score 2680; DB 1; Length 1166;		
Best Local Similarity 45.4%; Pred. No. 1.9e-102;		
Matches 574; Conservative 163; Mismatches 388; Indels 140; Gaps 29;		
Qy	4	AALCRWGLLLALPPGAAST-----QVC-----TOYIKANSKFGITETILLYOGCQVQGNLE 55
Db	8	AALLQ--LLLVLSIRCCSDPDPRKVCQGTSGNTQMTLDNHYLKWKKE-MYSGCNVLE 64
Qy	56	LTVLPNTASISFLQDQIEQGVYLVIAHNOVRQVLPQRLRVCTQLPEDNYALAVLDNG 115
Db	65	ITTQENQDLSFQSQIEQGVYLVIAWNEVSTIPLVNLRLIRGQNYEGNFTLVVMSNYQ 124
Qy	116	PLNNTPVTGASPGGLRELQRLSLTEILKGGVLIRQNPQLCYQDTILWKDIFHKNNQAL 175

Db 125 K-NPSSP--DVYQVGLKQLQLNLTEILSGGVKSHNPLLCNVETINWWDIVDKTSNPTM 181
QY 176 TLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACAGC-ARCKGQLPTDCCHEQC 234
Db 182 NLIPHAERQCKDCDHGCVNGSWAPGQHCQKFTKLLCAEQNRRRCRGPFPIDCCNEHC 241
QY 235 AAGCTGPKHSDCLACLHFNHSGICELHCHALVTYNTDTFESMPNBEGRYTFGASCVTACP 294
Db 242 AGCGTGPRATDCLACRDFNDGCTKCTPPKIYDIVSHQVVDNPNKITYTGAACVKCEP 301
QY 295 YNVLSTDVGSCTLVCPHMOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 354
Db 302 SNYVTE-CACVRSACMLEVD-ENGKRSCKPCDVCYKVCVDGIGISLSTIAVNSIN 359
QY 355 IQEAGACKIFGSLAFPLPESFGDPPASNTAPLQEOLOVFETLEETIGLYTISAMPDSL 414
Db 360 IRFSNCTKINGDIILNRNSFEGDPHYKIGTMDPEHLNLTITVKEITGYLVIMWPNMT 419
QY 415 DLSVFQNLQVIRGILHNGAYS-LTLQGLGISWGLRSRLRELSGLALHHTHLCHFVHT 473
Db 420 SLVSFQNLLEIRGRTTFSRGFSFVVQVVRHQLWGLRSLSKEYSAGNVILKNTLQLRYANT 479
QY 474 VPWDOLFRPHQALLHTANRPEDECVEGLACHOLCARGHCWGPPTQCVCNCSQFLRGOE 533
Db 480 INWRRLFRSEDGSIYDART-----ENQTCNNECEDGCW-PGFTMCVSLHVDGRGR 531
QY 534 CVECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVAR 593
Db 532 CVASCNLLGEPREAOQDRCVQCHQECVLQVDTSLTCYGPANCKSAHFQDQGCIPR 591
QY 594 CPSGVKPLSYNPIWKFPDEBAGACQPCPINCTHSCVDLDDKGCAPABQASPLTISVAV 653
Db 592 CPHGILGDGDTL-INKYADKMGQCQPCQNCCTQCGSGPGLSGCRGD-IVSHSLAVGLVS 649
QY 654 GILLVVLGVVGIILKREQKIRKVTWRLLQETELVEPLTPSGAMPNQAQWRLIKETE 713
Db 650 GLITITVALLIWLRRRIK-RKRTICLLOEXELVEPLTPSGAQPNAQFLRIKETE 708
QY 714 LRKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVS 773
Db 709 FKDRVLGSGAGTVYKGLWPDGENIRIPVAIKVIREATSPKVQEVLDAYVMASVDH 768
QY 774 PVSRLGLCLTSTVOLVTQMLPYGCLLDHVRNRRGLSGQDLNWCQIAKMSYLEVD 833
Db 769 PHVCRLLGLCLTSVQVLTQMLPYGCLLDYVROHQERICQWLLNWCQIAKGMVLEBR 828
QY 834 RLVHRDLAARNVLVSPNHVKITDGLARLLDIDETEHADGCKVPKIMMALESILRRF 893
Db 829 HLVRDLAARNVLLKNPNHVKITDGLSKLLTADEKEYQADGCKVPKIMMALESILQWY 888
QY 894 THQSDVWSYGVTVWELMTFGAKPYDGI PAEIPDLLEKGERLPQPPICITIDVYIMVKCW 953
Db 889 THQSDVWSYGVTVWELMTFGSPYDGI PAKEIASVLENGERLPQPPICITIEYIMILKCW 948
QY 954 MIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVD 1013
Db 949 MIDPSSRRPRFELVSEFSQWARDPSKYLVIQ--NLPSLSDRLFLSRLSSDD--DVVD 1003
QY 1014 ABEYLVPQGFPCPDPAAGAGWVHRHRSSTRSGGDLTLGLPESEEAAPSLAPSE 1073
Db 1004 ADEYLLPYKRI-----NRQS-----EPCIPPT 1026
QY 1074 GAGSDVFDGLGWAAGKGLQLPHTDPSPLQRYSEDPV-PLPSETDGYVAPLTCSPQE 1132
Db 1027 GH-----PVRENSITLRNISDPTQNALEKLDLGH-----E 1056
QY 1133 YNQPDVRRPQ-----PSPRE-----GPLP-AARPAGATLATERAKTSPGKNGVVKDVF 1179
Db 1057 YNQPGSETSSRLSDIYNPNYEDLTDGWPVSLSSQEAETNFSREYLTNTQNSL---PL 1113
QY 1180 AFGGAVENPEYLTPOGGAAPQHPHPPAFSPFDNLYYWDQDPPERGAPSTPKGTPTAEN 1239
Db 1114 VSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPAAEN 1151

QY 1240 PEYLG 1244
|||
Db 1152 LEYLG 1156

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.1-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor family
A:Reference number: A36223; MUID:90083234; PMID:2687875
A:Accession: A36223
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, G.J.; Todaro, G.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor-related protein
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'E', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PIDN:AAA35979.1; PID:g306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homology
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 35.6%; Score 2414; DB 2; Length 1342;
Best Local Similarity 40.6%; Pred. No. 1.6e-91;
Matches 531; Conservative 190; Mismatches 447; Indels 140; Gaps 33;

QY 5 ALCRWGLLLALLPPGAASTQVCTQVTKANSKFITIGITELLYQCGVQVQGNLELTYPNTAS 64
Db 27 AVCP-GTLNGLSVTGAENQYQTLV-----KLYECEVVMGNLEIVLTGNAD 73
QY 65 LSFLODIOEVOGYVLIHNVQVRQVPLQRLRIVRGQLFEDNYALAVLDNGDPLNNTPTV 124
Db 74 LSFQWIREVTGYVLVAMNEFTLPLNLRVVRGVQVYDGKFAIFVM-----LNYNT--- 125
QY 125 GASPGELRELQRLSLEILKGVLIQRPOLCYODTILWKDIFKKNQLALTLDITNRSR 184
Db 126 -NSHALRQLRLTQLEILSGGVYIEKNDKLCHMDTIDMRDIVDRD---ABIIVKDNGR 181
QY 185 ACHPCSPCKSGRCWGESSEDCQSLTRTVACAGC-ARCKGQLPTDCCHEQCAAGCTGPQH 243
Db 182 SCPPCHEVCKG-RWGPQSEDCQILTKITCAPQCNGCHCFGPNPNOCCHECACGSGGQD 240
QY 244 SDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNYLSTDVG 303
Db 241 TDCFACRHFNDGACVPRCPQPLVYNLTLFQLEPNPHTKYQYGVGVCAVCPHFV-VDQT 299
QY 304 SCTLVCPHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSANIQEAPGCKK 363
Db 300 SCVRACPPDKMEVD-KNGLKNCEPGGLCPKACEGTGSG--SRFQTVSSNIDGFVNCTK 356
QY 364 IFGSLAFPESEFGDGPASNTAPLOEQIQVFETLEETIGLYVISAWPDSLPLDSVFQNLQ 423
Db 357 ILGNLDFLITGLNGDPWHKIIPALDPEKUNVFRITVREITGYLNIQSWPPHMFNFSVSNLT 416
QY 424 VIRGRILHNGAYS-LTLQGLGISWGLRSRLRELSGLALIHHTHLCHFVHTVPMDQLFRN 482

Db 417 TTGSRSLYNRGFSLLIMKNLNTSLGFRSLKETISAGRIYISANRQLCYHSHLNTWKVLRG 476
QY 483 PHQALLHTA-NRPEDECYVEGLACHOLCARGHCWGPQTQVNCVSQFLRQECVBCRVL 541
Db 477 PTERLDIKHNRPRRDCVAEGKVCPLCSSGGCGPFGQCLSCRYSGGVCVTHCNFL 536
QY 542 QGLPRYVNAHCLPCHPECPQNGSVTCFGEADQCAHAYKDPFCFVARCPGSKVP 601
Db 537 NGEPREFAHEAECFCHPECPQMEGTATNGSGSDTCAQCAHPRDGHPCVSSCPHVLG- 595
QY 602 LSYMPITWKPDEGACQPCINCHSCVDLDDKCCPAEQRA-----SPLTSIISAVVGI 657
Db 596 -AKGPIYKYPDVQNECRCHENTCGCKGPELQDCLGQTLVLGKTHLTMALTVIAG--L 652
QY 658 VVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNQAQWRILKETLRK 716
Db 653 VVIFMVLGGTFLYWRGRRIONKRAMRYLERGESIEPLDPS-EKANKVLARIFKETLRK 711
QY 717 VKVLGSAFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKIELDEAYVMAGVSPYV 776
Db 712 LKVLGSGVFTVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLDAHI 771
QY 777 SRLGLCLTSTVQLTQMPYGLDHRNRCRLGSDLLNWCQIAKMSVLEDRVLV 836
Db 772 VRLGLGCPSSQLQVLYPLGLSLLDHVRQHRGALGPQLLNMVGQIAKMGTYLEHGVV 831
QY 837 HRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRFTHQ 896
Db 832 HRLAARNVLKSPSQVQVADFGVADLLPDDKQLLYSEAKTPIKWMALESIFGKYTHQ 891
QY 897 SDVWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWM 956
Db 892 SDVWSYGVTVWELMTFGASPYAGRLAEVPLLEKGERLAQPOICTIDVYIMVVKCWM 951
QY 957 SECRPFRELVSFESMARQPFVVIQNEDELGA---SPLDSTFRSLLEDDMDGLVD 1013
Db 952 ENIRPTEKELANETFMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKLEVELEPELD 1010
QY 1014 ABEYLVPQOGFFCPDPAPGAGGVHHRHSSSTRSGGDLTLGLEP-SEEAAPRSLAPS 1072
Db 1011 LDLDLEAED-----NLATTLGSLALSPLVGTILNRPGRSQSLSPS 1051
QY 1073 EGAGSDVFGDLGMAAGLQSLPTH-DPSLQRYSEDPTVPLP-----SETDGYVA-- 1123
Db 1052 SGY-MPMNQNGLSCQESAVSGSRCPRPVSLH-----PMRGLASSESSEGHVTS 1104
QY 1124 -----PLTCSQPE-----YVQPDVPRPQPSPREP-----LPAA 1154
Db 1105 EABLQEKVSMCRSRSRSPRGDSAYHSQRHSLTPTVTLSPPLGLEEDVNGYVMPDT 1164
QY 1155 RPAGATLERAKTLP-SGKNGV-----KDVFAFGGAVENPEYLTPOGGAAPOPHPPAF 1207
Db 1165 HLKGTSPSSREGTSLSSVGLSSVLGTEEDD-----EYFYMRRRHSP-PHPPRPS 1215
QY 1208 SPAFONLYWD-----QDPPERGAPSTFTKGTPTAENPEYL 1243
Db 1216 SLEELGYEYMDVGSLLSASLGSTQSCPLHPVIMPTAGTTPDEDYEM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternative names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision: 19-Apr-1996 #text_change: 13-Nov-1998
C:Accession: JC4387
R:Helleyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:g915389; PID:g915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GGT for residue
C:Comment: This protein is a functional heregulin receptor that transduces signals to ti
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)

Query Match 34.4%; Score 2333.5; DB 2; Length 1339;

Best Local Similarity 40.9%; Pred. No. 3.1e-88;
Matches 520; Conservative 169; Mismatches 422; Indels 161; Gaps 34;

QY 5 ALCRWGLLALLPPGAASQTQCTQYIKANSKEFIGITELLYQCGVVOGNLELTYPNTAS 64

Db 27 AVCP-GTLNGLSVTGDADNOYQFLY-----KLYEKEVVMGNLEIVLTCHNAD 73

QY 65 LSFLQDIQEVQSVYLIAHNQVRQVPLQRLRIYRGTLQFEDNYALAVLDNGDPLNNTPTV 124

Db 74 LSFLQWIREVTAYVLVAMNEFSLPLNLRVVRGTVQVYDGKFAIFVM-----LNYNT--- 125

QY 125 GASPGGLELQRLSITELKGVILQRPOLCYQDTILWKDIFHKNNQLALTLIDTNRSR 184

Db 126 -NSSHALRQLKFTQLTEILSGGVYIEKNDKLCHMDTIDMRDVRVR---GABEIVKNGGA 181

QY 185 ACHPCSPCKSGRCWGESSEDCQSILTRTVACGGC-ARCKGPLPTDCCHBOCAAGCTGPKH 243

Db 182 NCPPECHEVCKG-RCWGPDPDDCQILTKYICAPQNGRCFEGPNPQCCHECAGGSGPD 240

QY 244 SDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAPYNYLSTDVG 303

Db 241 TDCFACRRFNSGACVPRCPPELVNKLTFQLEPNPHTKYQYGGVCVASCNHFV-VDQT 299

QY 304 SCTLVCLPHNQVTAEDGTQCEKCKSPCARVCYGLGHEHLREVRVAVTSANTQEFAGCKK 363

Db 300 FCVRACPPDKMEVD-KHGLKMWCEPCGGLCPKACEGTSGS--SRYQTVDSSNIDGFCVNTK 356

QY 364 IFGSLAFIPESDGPDPASNTAPLOPELQVFPETLEEITGLYISAWPDSLPDSVFNQIU 423

Db 357 ILGNLDFLITGLNVDPWKIPALDPEKLNVRFTVREITGYLNIQSWPHMNFVSFSLT 416

QY 424 VIRGRILHNGAYS-LTLOGLGISWGLRSLRELGSGLALIIHNNHLCIFVHTVPWQLFRN 482

Db 417 TIGGRSLYNRGFSLLIMKNLNTSLGFRSLKETISAGRIYISANRQLCYHSHLNTWKVL 476

QY 483 PHQALLHTA-NRPEDECYVEGLACHOLCARGHCWGPQTQVNCVSQFLRQECVBCRVL 541

Db 477 PSEERLDIKYDRPLGECLEAGKVCPLCSSGGCGPFGQCLSCRYSGGVCVTHCNFL 536

QY 542 QGLPRYVNAHCLPCHPECPQNGSVTCFGEADQCAHAYKDPFCFVARCPGSKVP 601

Db 537 QGEPREFVHEAQCFCFCHPECLPMEGTSTYNGSGSDACARCAHFRDGHPCVNSCPHILG- 595

QY 602 LSYMPITWKPDEGACQPCINCHSC---VDLDDKGCPEAQRASPLTSIVSAVVGILLVY 659

Db 596 -AKGPIYKYPDAQNECRCHENTCGCGPELQDCLGQAEVLMKSPHLVIAVTVG--LAV 652

QY 660 VLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNQAQWRILKETLRKVK 718

Db 653 ILMLIGSFLYWRGRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFKETLRK 711

QY 719 VLGSAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKIELDEAYVMAGVSPYVSR 778

Db 712 VLGSVFTVYKGIWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLDAHI 771

QY 779 LLGICLTSTVQLTQMPYGLDHRNRCRLGSDLLNWCQIAKMSVLEDRVLVHR 838

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-250-264-12
Perfect score: 6801
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTEMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriapi.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6203	91.2	1259	6	018735 canis famil
2	3124	45.9	1208	11	Q9QX70
3	3095	45.5	1210	11	Q9EP98
4	2737	40.2	1165	13	Q9YH40
5	2696.5	39.6	1137	13	Q9W6F6
6	2287	33.6	1328	13	P79754
7	2002.5	29.4	1433	5	Q9BTH9
8	1782.5	26.2	419	4	Q9UK79
9	1739	25.6	367	11	Q8R2X1
10	1720	25.3	729	15	Q8G712
11	1718	25.3	567	15	Q8G714
12	1697.5	25.0	412	4	Q8WV00
13	1653.5	24.3	962	15	Q64895
14	1645	24.2	545	15	Q85468
15	1486.5	21.9	655	11	Q9WVFS
16	1470.5	21.6	643	11	Q9ERV6

17	1251	18.4	1193	5	Q9YIX8	Q9YIX8 ephydatia f
18	1167	17.2	1717	5	Q26566	Q26566 schistosoma
19	1158.5	17.0	1368	5	Q23821	Q23821 caenorhabdi
20	1108	16.3	527	13	Q90836	Q90836 gallus gall
21	981.5	14.4	478	11	Q9ESE0	Q9ESE0 rattus norv
22	973.5	14.3	599	13	Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	752.5	11.1	1362	13	Q9PVZ4	Q9PVZ4 xenopus lae
29	729	10.7	1671	5	Q9NJV5	Q9NJV5 biophalarai
30	702.5	10.3	1368	13	Q8UW85	Q8UW85 paralichthy
31	694	10.2	331	4	Q9BUD7	Q9BUD7 homo sapien
32	687	10.1	1418	13	Q93457	Q93457 scophthalmu
33	684.5	10.1	1369	13	Q8UW86	Q8UW86 paralichthy
34	669	9.8	1358	13	Q73798	Q73798 xenopus lae
35	652.5	9.6	1472	5	Q9U5A8	Q9U5A8 bombyx mori
36	648	9.5	1412	13	Q8UW84	Q8UW84 paralichthy
37	646	9.5	1245	13	Q9YGH8	Q9YGH8 scophthalmu
38	636.5	9.4	1418	13	Q8UW83	Q8UW83 paralichthy
39	631	9.3	149	6	Q9SG66	Q9SG66 oryctolagus
40	622	9.1	1371	11	Q9QVW4	Q9QVW4 rattus sp.
41	613.5	9.0	2144	5	Q9VD94	Q9VD94 drosophila
42	598	8.8	935	4	Q96L35	Q96L35 homo sapien
43	592	8.7	987	11	Q91YMO	Q91YMO mus musculu
44	589	8.7	987	11	Q99MR2	Q99MR2 mus musculu
45	587.5	8.6	1036	4	Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

018735	ID	018735	PRELIMINARY;	PRT;	1259 AA.
AC	018735;				
DT	01-JAN-1998	(TEMBLrel. 05, Created)			
DT	01-JAN-1998	(TEMBLrel. 05, Last sequence update)			
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)			
DE	Erbb-2.				
OS	Canis familiaris (Dog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.				
OX	NCBI_TaxID=9615;				
RN	[1]	SEQUENCE FROM N.A.			
RP					
RA	Yokota H.;				
RT	"cDNA cloning of erbb-2 from canine mammary gland."				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB008451; BAA23127.1;				
DR	HSSP; P11362; 1FGK.				
DR	InterPro; IPR002048; EF-hand.				
DR	InterPro; IPR000494; EGFR_L_domain.				
DR	InterPro; IPR000719; Euk_pkinase.				
DR	InterPro; IPR002174; Furin-like.				
DR	InterPro; IPR001245; Tyr_pkinase.				
DR	InterPro; IPR004019; YLP_motif.				
DR	Pfam; PF00757; Furin-like; 1.				
DR	Pfam; PF00069; pkinase; 1.				
DR	Pfam; PF01030; Recep_L_domain; 2.				
DR	Pfam; PF02757; YLP_2.				
DR	ProDom; PD000001; Euk_pkinase; 1.				
DR	SMART; SM00261; FU; 3.				
DR	SMART; SM00219; Tyrc; 1.				
DR	PROSITE; PS00018; EF_HAND; UNKNOWN 1.				
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.				
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.				
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.				
KW	ATP-binding; transferase; tyrosine-protein kinase.				
SQ	SEQUENCE 1259 AA; 137989 MW; E37364D49CADC46 CRC64;				

Db	15	LAALCAAG-----GALEBKVCQOQTSNRLTOLGTGPEDHFLSLQRMFNNCVVLGNLE	66
Qy	62	LTYLPTNASLFLQDIOEVOQGYVLIAHNQVQVQLRLRIVRGTQLFEDNYALAVLQNGD	121
Db	67	ITYQVQRYNDSLFLKTIQEVAGYVLIALTNTVERIPLENLIQIRGNALYENTYALAVLSN--	124
Qy	122	PLNNTTPVTGASPGGLRELQLRSITEILKGGVLTORNPOLCYQDTILWKDILFHKNQJAL	181
Db	125	-----YGFNKTGLRELPMRNLQETLIGAVRFSNNPILCNMETIQWRDITV-QDVFLSN	175
Qy	182	TLIDTNRS-RACHPCSPWCKSGRCSWGBESSDCQSLTRTVCAAGCA-RCKGPLPTDCCHEQ	239
Db	176	MSMDVQRHLTCCPKDPCSPNGSCWGGEENCQKLTKIIQAQCSRRCRGRSPSDCCHNQ	235
Qy	240	CAAGCTGPKSHSDCIACILFHNHSGICELHCPALVOYIKANSKFIGITELRYTFGASCVTAC	299
Db	236	CAAGCTGPRESDCLVCHFRDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCK	295
Qy	300	PYNLYSTDVGSCTIVCPLHNOEVTABDGTQRCCKSRPCARVCYGLGMEHUREVRAVTS	355
Db	296	PRNTVVTDHGSCVRACGPDYYEV--BEDGVSCKCKDGPCKVKCNIGIGTGFKDTLSINAT	354
Qy	360	NIQEFAGCKLFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEETIGYLVISAWPDSL	419
Db	355	NIKHFKYCTAISGDHLPLVAFKGDSTRTPPLDPRELEIUKTVEITGFLIIQAWPENW	414
Qy	420	PDLSVFQNLQVIRGRILHNGAYSITLQGLGTSWGLRSRLRBLGSLGALIIHHTHLCFVHT	479
Db	415	TDLHAFENLEIIRGTRKHQGFSLAVGLNITSLGLRSLKEISDGDVVISGRNLCYANT	474
Qy	480	VPWQLFRNPHQALLHTANRPDEDCVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQOE	539
Db	475	INWFKLFGTPNQKTKIMNRAEKDCKTANHVNCNPLCSEGCWGPEDTDCVSCQVSRGRE	534
Qy	540	CVESCRVLQGLPREVYNARHCLPCHBPQOPQNGSVTCFGRPADOCVACHYKDPFCVAR	599
Db	535	CVDKCNILGEGPREFVENSECIIQCHPECLPOTMNICTGRGPDNCIIKAAHYVDGPHCVKT	594
Qy	600	CPSGVKPDLSTYMPWKPDDEBAGACQPCPINCTHSCVDLDDKGCFAEQRASP-LRFSIVSAV	658
Db	595	CPSGIMGENNTL-VKKADANNVCHLCHANCTYGCAGPLKGC--QOPEGPKISATGI	651
Qy	659	VGILLVVVLGVVFGI-LIKRQOKIRKYTHRRLLQETVELPELTPSGAMPNQAOMLIKE	717
Db	652	VGGLLFIWV-VALGIGLFMRRLQVRKTRLLRLLOERELVELPELTPSGEAPNQAHLRLKE	710
Qy	718	TELKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSFKANKEILDEAYMAGV	777
Db	711	TEFKKIKVLGSGAGTGYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYMAGV	770
Qy	778	GSPVYSRLIGICLSTVOLTPQLMPYGCGLLDHVRENRLGSGDLLNWCQIAKMSVLE	837
Db	771	DNPHVCRLLGICLSTVOLITQLMYPYGCGLLDVYREHKDNTGSOYLLNWCQIAKMNYLE	830
Qy	838	DVRLVHRDLAARNLVKSPNHVKITDIFGLARLLDIDETEHADGGKVPKIKWMALESILRR	897
Db	831	DRRLVHRDLAARNLVKTPQHKITDIFGLAKLGAEBKEYHAEGKVPKIKWMALESILHR	890
Qy	898	RPTHQSDVMSYGVTVWELMTFGAKPYDGIPIARETIDLLEKGERLPPQPICTIDYVIMVVK	957
Db	891	IYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSLILEKGERLPPQPICTIDYVIMVVK	950
Qy	958	CWMTDSECRPRFRELVSFEFSMARDPORFVVIQ-NEDLGPASPLDSTFYRSLLEDODMDG	1016
Db	951	CWMLDASGRKRELIEUFSKWARDPQRYLVIQGDERMHLPSPTDSNYPYRALMEEDMED	1010
Qy	1017	LVDAAEYLVPOQGFPCPDPAFGMVGWHRRHSRSTSGGGDLTLGLEPSEEPSEAPRPLA	1076
Db	1011	VWDAEYLVPOQGF-----NSPST-----SRPPLL	1036
Qy	1077	PSEGAGSDVFDGDLGMGAACKQLSLPHDPSPLQRYSEDPVTPLPSET--DGYVAPITCS	1134
Db	1037	SSLGANSN----SSTVACINRNGSCRKVEDAFLQYSSDPTSVLTENIDIDTFL-----	1086

[illegible]


```
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA, 134840 MW, 62CD021C9DE32E18 CRC64;

Query Match 45.58; Score 3095; DB 11; Length 1210;
Best Local Similarity 49.48; Pred. No. 3.8e-224;
Matches 628; Conservative 166; Mismatches 366; Indels 110; Gaps 23;

QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPEHDLMLRLHYOCQVQVQGNLELTLYPTN 68
DB 11 LLLALLPPGAA--STQVCTGTDMLRLPASPEHDLMLRLHYOCQVQVQGNLELTLYPTN 68
QY 14 LLLTALCAAGALBEKKVCOQTSNRLTQGTFFDHFLLSLQMYNCEVVLGNLEITYVORN 73
DB 14 LLLTALCAAGALBEKKVCOQTSNRLTQGTFFDHFLLSLQMYNCEVVLGNLEITYVORN 73
QY 69 ASLSFLODIQEVGYVLIANQVQVPLQRLRIVRGQLFEDNYVALAVLDNGPLNNTTP 128
DB 69 ASLSFLODIQEVGYVLIANQVQVPLQRLRIVRGQLFEDNYVALAVLDNGPLNNTTP 128
QY 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLIQIRGNALYENTYVALAILSN----- 124
DB 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLIQIRGNALYENTYVALAILSN----- 124
QY 129 VTGASPGELRELQRLSITELIKGVLIQRLPQCYDTILWKDI-----FHNQLALTLI 184
DB 129 VTGASPGELRELQRLSITELIKGVLIQRLPQCYDTILWKDI-----FHNQLALTLI 184
QY 125 -YGTNRTPGLRELPMRNLQELILGAVRFSNPILCNMDTIQWRDIVQNVFMSNMDL--- 180
DB 125 -YGTNRTPGLRELPMRNLQELILGAVRFSNPILCNMDTIQWRDIVQNVFMSNMDL--- 180
QY 185 DTRNSRACHPCSPMCKSGRCWGESSEDCQSLTRVCAGGCA-RCKGPLPTDCCHQCAAG 243
DB 185 DTRNSRACHPCSPMCKSGRCWGESSEDCQSLTRVCAGGCA-RCKGPLPTDCCHQCAAG 243
QY 181 -QSHPSCKPCDSCPNWSCWGGGEENCQKLTIIICAOQCSHRGCRSPSDCCNQCAG 239
DB 181 -QSHPSCKPCDSCPNWSCWGGGEENCQKLTIIICAOQCSHRGCRSPSDCCNQCAG 239
QY 244 CTGPKHSDCLACLFHFNHSGICELHCPALVQYIIRKANSKFIGITELRYTFGASCVTACP 303
DB 244 CTGPKHSDCLACLFHFNHSGICELHCPALVQYIIRKANSKFIGITELRYTFGASCVTACP 303
QY 240 CTGPRESDCLVCKQFQDEATCKDTPPLMLNPTTYQMDVNPBGKYSFGATCVKCKPRNY 299
DB 240 CTGPRESDCLVCKQFQDEATCKDTPPLMLNPTTYQMDVNPBGKYSFGATCVKCKPRNY 299
QY 304 LSTDVSGCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRATVANSIQE 363
DB 304 LSTDVSGCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRATVANSIQE 363
QY 300 VVTDHSGCVACRACPDYVEV-EEDGIRKCKCDGCRKVCNGIGIGEFKDFLTSINATIKH 358
DB 300 VVTDHSGCVACRACPDYVEV-EEDGIRKCKCDGCRKVCNGIGIGEFKDFLTSINATIKH 358
QY 364 FAGCKIFGSLAFPEFDGDPASNTAPLOEQLVPELTLEETIGYLIISAWPDSLPDLS 423
DB 364 FAGCKIFGSLAFPEFDGDPASNTAPLOEQLVPELTLEETIGYLIISAWPDSLPDLS 423
QY 359 FKYCTAISGDHLPLVAFKSDSFTRTPPLDRELEILTKVETITGFLIIQAWPDWTDLH 418
DB 359 FKYCTAISGDHLPLVAFKSDSFTRTPPLDRELEILTKVETITGFLIIQAWPDWTDLH 418
QY 424 VFQNLQVIRGIRLHNGAYSITLQGLGISWGLSLRLSLGLSLGLIHHNTHLCFVHTVPWD 483
DB 424 VFQNLQVIRGIRLHNGAYSITLQGLGISWGLSLRLSLGLSLGLIHHNTHLCFVHTVPWD 483
QY 419 AFENLEIIRKTRKHQGFQSLAVVGLNITSLGSLSLKEISDGVIIISGNRLCYANTINWK 478
DB 419 AFENLEIIRKTRKHQGFQSLAVVGLNITSLGSLSLKEISDGVIIISGNRLCYANTINWK 478
QY 484 QLFNPHQALLHTANRDEDCVGEGLACHOLCARGHCWGPGTQCVCNCSOFLRGQECVEE 543
DB 484 QLFNPHQALLHTANRDEDCVGEGLACHOLCARGHCWGPGTQCVCNCSOFLRGQECVEE 543
QY 479 KLFGTNPQTKIMNNAEKDCKAVNHCNPLCSEGCWGPEDPCVSCQNVSRGECVEK 538
DB 479 KLFGTNPQTKIMNNAEKDCKAVNHCNPLCSEGCWGPEDPCVSCQNVSRGECVEK 538
QY 544 CRVLQGLPREYVNAHCLPCHEPCQONGSVTCFGEADOCVCAHYKDPFPCVAPCPG 603
DB 544 CRVLQGLPREYVNAHCLPCHEPCQONGSVTCFGEADOCVCAHYKDPFPCVAPCPG 603
QY 539 CNTLEGEPRFVENSECIIQCHPECLPOAMNITCTGRGPDNCIQCAHYIDGPHCVKTCAG 598
DB 539 CNTLEGEPRFVENSECIIQCHPECLPOAMNITCTGRGPDNCIQCAHYIDGPHCVKTCAG 598
QY 604 VKPDLSTYMPKPPDEEGACQPCPINCTHSCVDLDDKCPAQASPLTSIVSAVVGILL 663
DB 604 VKPDLSTYMPKPPDEEGACQPCPINCTHSCVDLDDKCPAQASPLTSIVSAVVGILL 663
QY 599 IMGENNTL-VKQYADANNVCHLCHANCTYGCAGPGLQCEVWPSPGKIPSIATGIVGGLL 657
DB 599 IMGENNTL-VKQYADANNVCHLCHANCTYGCAGPGLQCEVWPSPGKIPSIATGIVGGLL 657
QY 664 VVVLGVVFGI-LIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETLRK 722
DB 664 VVVLGVVFGI-LIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRILKETLRK 722
QY 658 FIWV-VALGILFLMFRRHVIRKTRLLQLQERELVEPLTPSGEAPNOAHLRIKETFEFK 716
DB 658 FIWV-VALGILFLMFRRHVIRKTRLLQLQERELVEPLTPSGEAPNOAHLRIKETFEFK 716
QY 723 VKVLGSAFTGYKGIWIPGENVKIIPVAIKVLRNTPSKANKEILDEAVYVAGVGSPPV 782
DB 723 VKVLGSAFTGYKGIWIPGENVKIIPVAIKVLRNTPSKANKEILDEAVYVAGVGSPPV 782
QY 717 IKVLGSAFTGYKGIWIPGEKVKIIPVAIKELREATSPKANKEILDEAVYVMSVDNPHV 776
DB 717 IKVLGSAFTGYKGIWIPGEKVKIIPVAIKELREATSPKANKEILDEAVYVMSVDNPHV 776
QY 783 SRLIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSLQDILLNMCQIAKMSYLEDLVRL 842
DB 783 SRLIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSLQDILLNMCQIAKMSYLEDLVRL 842
QY 777 CRLIGICLTSTVQLITLQMPYGCILLDVREHKDNIGSQYLLNMCVQIAKGMVLEDRRLV 836
DB 777 CRLIGICLTSTVQLITLQMPYGCILLDVREHKDNIGSQYLLNMCVQIAKGMVLEDRRLV 836
QY 843 HRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRPETHQ 902
DB 843 HRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRPETHQ 902
QY 837 HRDLAARNLVKTPQHVKITDFGLAKLLGAEKYEYHAEGKVPKIKWMALESILRIYTHQ 896
DB 837 HRDLAARNLVKTPQHVKITDFGLAKLLGAEKYEYHAEGKVPKIKWMALESILRIYTHQ 896
QY 903 SDVMSYGVTVWELMTFGAKVDGIPAREIPDLLEKGERLPQPPCTIDVYMWVKWCMID 962
DB 903 SDVMSYGVTVWELMTFGAKVDGIPAREIPDLLEKGERLPQPPCTIDVYMWVKWCMID 962
QY 897 SDVMSYGVTVWELMTFGSKYDGIIPASDISILEKGERLPQPPCTIDVYMWVKWCMID 956
DB 897 SDVMSYGVTVWELMTFGSKYDGIIPASDISILEKGERLPQPPCTIDVYMWVKWCMID 956
QY 963 SECRPRFRELVSFSRWARDQRFVITQ-NEDLGASPLDSTFYRSILLEDDDDGDLVDAE 1021
DB 963 SECRPRFRELVSFSRWARDQRFVITQ-NEDLGASPLDSTFYRSILLEDDDDGDLVDAE 1021
```



```
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWLGRLSLRELGLSLALIHNTLHLCFVHTV 480
Db 404 DLSVFSLLTIQGRSLFKRFLSMVMRITPTLTSLGLSLREISDVGSVYISQNAHLCHYHTV 463
Qy 481 PWDQLFRNPH-QALLHTANRPEDBCVGEGLACHOLCARGHCWGPGTQCVCNCSQFLRGQE 539
Db 464 NWTQLFGRSVRANSLNSNRPMAECVADGRVCDPLCSDSGCWGPGPQCLSCRYNHSHT 523
Qy 540 CVBECRVLOGLPREYVNAH-RCLPCHPEQOPNGSVTCFGEADQCACAHYKDDPPPCVA 598
Db 524 CVAGCHFNISGIPREFAGLNGVCVACHPECKPQTKASCTGPGADECMACFKFRDGPYCMS 583
Qy 599 RCPGKVPDLISYMPIWKFPEDEGACQPCINCHTSCVDLDDKGCAPQORASPLTSIVSAV 658
Db 584 SCPAGVN-DEKGLIFKPNREGHCEPHQNCQTCGCGPGLNDC---LEAARLTSSGQI 639
Qy 659 VGILLVVVLGVVF-----GILIKRQOKIRKYVMRLLQETVELVBLTPSGAMPNQAQ 711
Db 640 TGIALGVPAGLIFCLVLFGLMLYHRLATRRKAMRYLESSESFEPLGP-GEKGTKVH 698
Qy 712 MRILKETELRVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 771
Db 699 ARILKPSDLRKIKPLGSGVGTGSKGFWIPEGETVKIPVAIKTIQDSSGRQTFTETDHL 758
Qy 772 YWAGVGSPIVSRLLGLICTSTVQLVTQMLPYGCLLDHVRENRLGSLQDLNWCWQIAK 831
Db 759 LMSGSLDHPYIVRLGICPGTCLQLVTLQSLSSHSLLEHIRHQTSLDPPQRLNWCWQIAK 818
Qy 832 GMSYLEDVRLVDRDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMAL 891
Db 819 GMYLEHRVHVHKNLAARNILLKNDYQVSDYGVADLLYPDDKKYVYSETKTIKIMMAL 878
Qy 892 ESILRRFRTHQSDVMSYGVTVWELMTFGKPYDGIPIAREIPDLLEKGERLPQPPICTIDV 951
Db 879 ESILFRYTHQSDVMSYGVTVWEMMSFGAPYASQVQPEVPSVLEKGERLSQAICTIDV 938
Qy 952 YIMVVKWMDISCRPFRELVEFSFMRADDPQRFVVIQNEGLGASPLDSTFYRSLLED 1011
Db 939 YIMVVKWMDENIRPTFKELASDFTMRADDPYRLVIRMEG-----ED 982
Qy 1012 DMGDLVDABEYLVPQGGFCPPAPGAGGMVHRHRSSTRSGGDLTLGLSPSREEAP 1071
Db 983 SGHGEPI-----RGSR-----GLLEADLEDEEE-- 1008
Qy 1072 RSLAPSEGAGSDVFDGLQMG---AAKGLQSLPTHDPSPQLQ-----RYSEDPTV 1118
Db 1009 -----GLGDRFATPSLQPSWSNSTPSQINSYVMVTLQAYD----- 1044
Qy 1119 PLSEDTGYVAPLTCSPQ- EYVQ-----PDVRPQPPSPREGPL--PAAR 1161
Db 1045 -FAVSQGGHIGYLPMSPPVDITIRLWYQSRSLSSVRTLPDRSAFRSSREABELCEDGAQ 1103
Qy 1162 PAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPOGGAAPQHPHPPAFSPADNLY 1221
Db 1104 CAGIFVR-----FGERGN-----PQGG----- 1122
Qy 1222 YWQDPPPERGAPPTFKGTPTAENPE 1247
Db 1123 --QQRKLSTASSPSSPKTWADEDE 1146
```

RESULT 7

```
Q9BIH9 ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
```

```
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
KW NON_TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D8967724F07 CRC64;
```

```
Query Match 29.4%; Score 2002.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred. No. 11e-141;
Matches 469; Conservative 196; Mismatches 397; Indels 385; Gaps 36;
```

```
Qy 26 CTGDMKLRPASPEETHLDMLRHLRYOGCQVVOGNELELYPTNASISFLQDIQEOVGYYL 85
Db 1 CIGTNGMSVPANREYHYKNLRDRYTNCTYVDNGLEITWIQNTLDLNFLOHIREVTGYVL 60
Qy 86 IAHNVQVQPLQRLRIVRGTLF-----EDNVALAVLDNGDPLNNTTPTVGTASPGGLREL 140
Db 61 ISLYDLQPVLPLQIIRGRTTFLKNKEEAYGLFV-----SFSHMTL 104
Qy 141 QLRSLTEILKGVLIQORNPOLCVQDTILWKDI-FHKNNOLALTLDITNRSRACHPCSPMC 199
Db 105 ELPALNDILGSGVGFNNVNLCHKMSINWEEILLAPQTSWQYTFNPFSPERVCPPCHPSC 164
Qy 200 KGSRCWGESSEDCQSILTRTVACAGCA--RCKGPLPTDCCHCCQAAGCTGPKHSDCLACLH 257
Db 165 EVG-CWGECAHNCQRFSLKNCSPQCSQGRFCFGKPRECCHLFCAGCTGTQSDCLACKN 223
Qy 258 FNHSGICELHCBALVOYIKANSKFIGITELRYTFGASCVTACPNYVLTDSVGSCTLVCP 317
Db 224 FYDDGVCKOECPMQIYNPTNYFWBPNDGKYAYGATCVRKCP-BHLLKDNAGACVRKCP 282
Qy 318 HNOEVTAEQGTORCEKSKPCARVCYGLGMEHLREVRVTSANIQEFACCKIFGSLAFL 377
Db 283 GKMPQNSE-----CVFKGVCPKTCPEGIVH-----SDNIGNKDCITIEGSELEL 329
Qy 378 PESFDGDPASNT-----APLQEQLOVFFETLEBEITGYLYISAWPDSLPDLSPFQNIQ 429
Db 330 DOSFDGQQVYTFNFGPRYIKIDPDRLEVTFVKEITGFINIQAHHPNFTLLNYFRNLE 389
Qy 430 VTRGRIILHNGAY-SLTQGLGISWLGRLSRLRELGLSLALIHNTLHLCFVHTVWDQLPFN 488
Db 390 VVGRQLKENLPASVYIVKTSLSLELSLKRNVSGSIVILENSDLCFVEDIDWSEIKKS 449
Qy 489 PHQALLHTANRPEDBCVGEGLACHOLCARGHCWGPGTQCVCNCSQFLRGQECVECRVLQ 548
Db 450 SDHEVMVQKNRNATTECHBEEMECBQCSKAGCWGKPGBOCLECKNVKYKGLDCK---- 506
```

```
QY 549 GLPREY-VNARHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPFCVRCVP----- 601
DB 507 SLRPLYSVDSKTCGDCHQCKD-----FCYGNEDNCGSCMNVDGGRFCVACPTTKHAM 561
QY 602 -----SGVKPDLSYMPYIWKFPD----- 618
DB 562 NGTCINCHKTCVGRGRPRDTIAPDGCISCDAKIIGSDAKIERCLMKDSCPDGYSDYVL 621
QY 619 -EG----- 621
DB 622 QEEGPKQLSGKAVCRKCHPRCKKCTGYFHEQFCQECTGYKKGEOCEDCPQDFYANEE 691
QY 622 --ACQCPINCT-----HSCVDL-----DD-----KGCPEAQ----- 646
DB 682 TRICLPCHQCRGCHGLGDHHECRNLKLFEGDPYDNATFTCVSNCPASHPKYRPOEA 741
QY 647 -----RASPLTSIVSAVGILVVLGVVFGI---LKRROQKIRKYTM 687
DB 742 GKIGPYCSADSMQSLRIEPTQVKIVGSGMALILLCVVFGIAFVLSRHKNKDAVKM 801
QY 688 RLLOETELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVK 747
DB 802 TMLAGCEDSELRPSNVGNLTKRIKEAIRRGVGLWGAFGRVFKGVMPGESVK 861
QY 748 IPVAIKVIRENTSPKANKEILDAYVMAGVSPVSRLLIGLCTSTVQLVTOLMPYGCCLL 807
DB 862 IPVAIKVIMEMSGSESEKFELEAYINASVEHNLKLLAVCHTSQMWLLITQLMPJGCLL 921
QY 808 DHVRENRGRIGSDQLNWCQIAKGMYSLEDRVLRVHRDLAARNVLKSPNHVKITPFLA 867
DB 922 DYVRNNKDKIGSKALLNWSQIARGMAYLEERLVRHDLAARNVLQTPSCVKITVFLA 981
QY 868 RLIDIDETEYHAGGKVPKIMMALESILRRRFTHOSDVWYGVTVWELMTFGAKPYDGP 927
DB 982 KLLDFDSDEYRAGGKMPKWLALCIRHVFVTSKSDVWAFGTTIWELLTYGARPVENVP 1041
QY 928 AREIPDLEGERLPQPICTIDVYIMVWKCWIDSECRPRFRELVSFSEFMRARDPORFV 987
DB 1042 AKDVELIEIGHKLPPQDICSLDVYCIILSCWLDADARTFQLAETFAEKARDPGRYL 1101
QY 988 VIQNEGLGASPLDSTPYRSLLEDDMGDLV----- 1018
DB 1102 MI-----PGDKFMRLEPSYTNQDEKDLIRTLAPVMAAAAAAAAAAGASNVDPSTIA 1152
QY 1019 DAREYLPVQOGFCPCDAPGAGGVHRRSSSTRSGGDLTLCLEPSEBAPRS----- 1073
DB 1153 ETEYLOPKTRPSIMLPESA-----GVGGIR-----VEPS-DEMPKSLRYCK 1188
QY 1074 -PLAP---SEGAGSDVPDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVA 1129
DB 1189 DPLKPDDETGHGKEV-----GVGGIR-----LNLPLDEDDYLM 1222
QY 1130 PLTCSPQEVYVNDVPRQPPSPREGPLPAARPAAGATLERAKTLPSPKGVGVKDVFAFG 1189
DB 1223 P-TCQSQ---NQS-----TPG-----YMDLIGVPA 1243
QY 1190 AVENPEYL-----TPQGAAPQHPBPAPSPAFDNLVYWDQDPPERGAPPSIFKGT 1240
DB 1244 SVNDPEYLMGSGTQIAGLAGSMG--PHTPP-----PNTPNMG 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287
```

RESULT 8

```
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
```

```
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 26.2%; Score 1782.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 6.6e-126;
Matches 340; Conservative 7; Mismatches 40; Indels 7; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTMKRLRPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLLALLPPGAASQVCTGTMKRLRPASPETHLDMLRHLYQGCVVQGNL 60
QY 61 EUTYLPNLSIFLODIQSVQGVLIHAHQVQPLQRLRIVRGTLFEDNYALAVLDNG 120
DB 61 EUTYLPNLSIFLODIQSVQGVLIHAHQVQPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQLRLSTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTPTVTGASPGGLRELQLRLSTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
QY 241 AGCTGPKHSDCLACLNHNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300
DB 241 AGCTGPKHSDCLACLNHNHSGICELHCPALVQYIKANSKFIGITELRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTIVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
DB 301 YNYLSTDVGSCTIVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
QY 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOP 394
DB 356 LRMQPG--PAHPVLSFLRPSDWLVSFAFYSLLPLAP 387

RESULT 9
Q9R2X1 ID Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

DR ENBL; BC027080; AAH27080.1; -;
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 04163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1e-122;
Matches 32; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPQPPICT 60

Qy 949 IDVYIMVWKWMDSECRPRRELVSFSESRMARDPQRFVIVIONEDLGPASPLDSTFVRSL 1008
Db 61 IDVYIMVWKWMDSECRPRRELVSFSESRMARDPQRFVIVIONEDLGPSPMDSTFVRSL 120

Qy 1009 LEDDMDGLVDABEYLVPOQGFCDPAPGAGGVHRRHSSSRSGGGDLTLGLEPSEE 1068
Db 121 LEDDMDGLVDABEYLVPOQGFSPDPAFGTGTSTHRRHSSSRSGGGDLTLGLEPSEE 180

Qy 1069 EAPRSLAPSEGAGSDVFDGLGMAAKGLQSLFTHDPSPLOQRYSEDTVPLPSETDGYV 1128
Db 181 EPPRSLAPSEGAGSDVFDGLAVGVTGKGLQSLSPHDLSPLOQRYSEDTVPLPETDGYV 240

Qy 1129 APLTCSQPEYVNPQDVVRPPSPREGPLPAAPAGATLERAKTLSPGKGVVVDVPAFG 1188
Db 241 APLACSPQPEYVNPQDVVRPPSPREGPLPAAPAGATLERPKTLSPGKGVVVDVPAFG 300

Qy 1189 GAVENPEYLPQCGAARPPHPPAFSPAFONLYWQDPPERGAPSTFGTPTAENPEY 1248
Db 301 GAVENPEYLPQCGAARPPHPPAFSPAFONLYWQDPPERGAPSTFGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712; MEDLINE=94203659; PubMed=8152791;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR ENBL; S69372; AAC60725.1; -;
DR HSP; P03322; JAK6.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match 25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 7.7e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

Qy 569 PQNGSVTCGPEADQCVACAHYKDPFCVACRCPGKVPDLSPYPIKFPDDEGACQPCPI 628
Db 141 PEETATPKTGP--DHCMKCAHFIDGPHCVACRCPGVLGENDTL-VWKYADANAVCQLCHP 197

Qy 629 NCTHSCVDLDDKCPAEQASPLTISVAVV-GILLVWVVLGVVFGILIKRRQOKIRKYTM 687
Db 198 NCTRGCKGFLGECPP--NGSKTPSTAAAGVVGGLLCLVVGVLGILYLR--HIVKRTLL 253

Qy 688 RLLQETELVEPLTPPGAMPNQAMRLKETELRKVKVLGSGAFGVYKGIWIIPDGENVK 747
Db 254 RLLQERELVEPLTPSGEAPNQAHRLKETETPKVKVLGSGAFGVYKGLWIPGEKVK 313

Qy 748 IPVAIKVLRENTSPKANKETLDEAYVMAGVGVSPYVSRLLIGICLTSTVQLVTLMPYGCLL 807
Db 314 IPVAIKELRENTSPKANKETLDEAYVMASVDNPRVRLIGICLTSTVQLTQLMPYGCLL 373

Qy 808 DHVRENRRGLGSODLLNWCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLA 867
Db 374 DYIREHKDNIQSQYLLNWCQIAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLA 433

Qy 868 RLLDIDETEHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 927
Db 434 KLLGADEKETHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 493

Qy 928 AREIPDLLEKGERLPQPICTIDVYIMVWKWMDSECRPRRELVSFSESRMARDPQRFV 987
Db 494 ASEISSVLKGERLPQPICTIDVYIMVWKWMDSDSRPKRELIAEFSKMARDPQRYL 553

Qy 988 VIO-NEDLGPASPLDSTFYRSILLEDDMDGLVDABEYLVPOQGFCDPAPGAGGVVHR 1046
Db 554 VIQDERMHLPSPTDSKFTYRLMEEDMEDIADAEYLVPHQGF-----598

Qy 1047 HRSSSTRSGGDLTLGLEPSEBEAPRSL-----APSEGAGSDVFDGLGMAAKGLQSL 1101
Db 599 -NSPST-----SRTPLSLSLSATSNNSATNCID-----RNGQGH 631

Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNPQDVVRPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGTFNLEESIDGDL-----PAPEYVNG--LMPKKPS-----675

Qy 1160 ARPAGATLERAKTLSPGKGVVVDV-----APGGAIVENPEYL 1197
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRVQNSHSTAVDNPYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714; MEDLINE=94203659; PubMed=8152791;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR ENBL; S69372; AAC60727.1; -;
DR HSP; P11362; lFGK.
DR InterPro; IPR000719; Euk_pkinase.

```
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CA7ADF725E1 CRC64;

Query Match 25.3%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 7.5e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPSADQVCAHYKDPFCVARGSPGVKPLSYNPIWKFPDREGACQPCINCHTSCVDL 637
DB 1 GP--DHCMKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 57
QY 638 DKGCPAEQRASPLTSTVSAVV-GILLVVVLGVVFGILIKRRQKIRKYMRLLOTEL 696
DB 58 GLEGCP---NGSKTPSIAAGVGGLLCLVVGIGLYLRRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNQAQRIKTELKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHLRIKETEFKVKVLGSGAFGVYKGLWIPGEKVKIPIVAIKEL 173
QY 757 ENTSPKANKIILDEAYVMAGVSGPYVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRNRR 816
DB 174 EATSPKANKIILDEAYVMASVDNPRVCRLLIGICLTSTVQLITQMPYGCILLDIYREHKON 233
QY 817 LGSODLLNWCQITAKGMSYLEDLVRLDIAARNVLKSPNHVKITDFGLARLLDDETE 876
DB 234 IGSQYLLNWCQITAKGMYLEERLVRDLAARNVLKTPQHKITDFGLAKLLGADEKE 293
QY 877 YHADGGKVPKIMWALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
DB 294 YHAEGGKVPKIMWALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIPDLLE 936
QY 937 KGERLPOPPCTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPOPPCTIDVYIMVKCWMIDADSRPKFRELIAEFSKWARDPPRYLVITQGERMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRRHSSSRSG 1055
DB 414 LPSTUSKFRYLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLTGLPSESEAPRSPL-----APSEGAGSDVFDGLGMGAAGLQSLPHDPSPLQ 1110
DB 455 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFVQ 491
QY 1111 RYSEDFTVPLPSET--DGYVAPITCSQPYVYNQPDVRPQPPSPREGPLPAARPAGATLE 1168
DB 492 RYSSDPTGNFLESIDGFL-----PAPEYVQ--LMPKKPS----- 526
QY 1169 RAKTLPFGKNGVKQVF-----AFGGAVENPEYL 1197
DB 527 -----TAMVQNYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P36559.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.

InterPro; IPR001245; Tyr_kinase.
Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
growth.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 1.6e-119;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESTLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 948
DB 1 MALESTLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICT 60
QY 949 IDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGASPLDSTFYRSL 1008
DB 61 IDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLGASPLDSTFYRSL 120
QY 1009 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRRHSSSRSGGDLTLGLPSPSE 1068
DB 121 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGGMVHRRHSSSRSGGDLTLGLPSPSE 180
QY 1069 EAPRSLAPSEAGSDVFDGLGMGAAGLQSLPHDPSPLQRYSEDFTVPLPSETDGYV 1128
DB 181 EAPRSLAPSEAGSDVFDGLGMGAAGLQSLPHDPSPLQRYSEDFTVPLPSETDGYV 240
QY 1129 APLTCSQPYVYNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPKNGVVKDYVAFG 1188
DB 241 APLTCSQPYVYNQPDVRPQPPSPREGPLPAARPAGATLERAKTLSPKNGVVKDYVAFG 300
QY 1189 GAVENPEYLTPOGGAAPQ-----HPPA---FSPAFDNL 1220
DB 301 GAVENPEYLTPOGGAALPTLLPSAQPSSTTSITGTRTHQSGGLHPAPSKGHLRQRTQST 360
QY 1221 YVWD-QDPPER-----GAPPTFGKTPTAEN 1245
DB 361 WWTQCEPEGVRRSPDVSSSREGSLTSAGIKRWEGPPTTSRGTCCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag v-erb-A v-erb-B protein.
DE Gag v-erb-A v-erb-B.
GN GAG v-erb-A v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1959616;
RA Bruskini A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
transforming potential of the oncogene v-erb-B.";
RT
```

```

Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroidfinger.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Nuclease; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1.2e-115;
Matches 350; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHPEQC 568
DB 354 IEKQESYLAFEHYINRYKHPIHFWSKILMKVADLRMTGAYHASFHMKVCEPTELS 413
QY 569 PQNGSVTCFGEADQCACAHYKDPFCVACRPSGVKFDLSYMPIWKFDPDEGACQPCPI 628
DB 414 PQE-----VGP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 465
QY 629 NCTHSCVDLDDKCGPAQRASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTM 687
DB 466 NCTRGCKGPLEGCP---NGSKTPSIAAGVVGGLCLVVGIGLGLYLR--HIVRKRTL 521
QY 688 RLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVK 747
DB 522 RLLQERELVEPLTPSGEAPNQAHRLILKETEFKKVKVLSGAGFTVYKGLWIPGEKVT 581
QY 748 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYVRLGLIGLCLTSTVQLTQLMFYGCLL 807
DB 582 IPVAIKELREATSPKANKEILDEAYVMASVDNPHVCHLLGICLTSTVQLTQLMFYGCLL 641
QY 808 DHVRENKRLGSDQLLNCWQIAKMSYLEDLVRLVHRLDAARNVLKSPNHNKIPFGLA 867
DB 642 DYIREHNDNIGSQVLLNWCQIAKGMNYSERHVRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLIDIDETEYHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 927
DB 702 KQLGADKEEYHAGGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 761
QY 928 ARBIPDLLEKGERLPQPICTIDVYIMVWKCMIMDSECRPRFRELVSFSEFMRMARDPORFV 987
DB 762 ASIISSVLEKGERLPQPICTIDVYIMVWKCMWSGADSRPKFRELIAEFSEFMRMARDPRYL 821
QY 988 VIQ-NEDLGASPLDSTFYFSLLEDDMDGLVDAEYLVPOQGFPCPDPAFGAGVMVHR 1046
DB 822 VIQGDERMHLPSPTDSKFSYRTLMEEDMEDIVDAEYLVPHQGF-----866
QY 1047 HRSSTFSGGDDITLGLPESEERAPSPAPSGAGSDVDGDLGWGAAGLQSLPHTDP 1106
DB 867 -NSPST-----SRTPLLSLSATSN-----NSATKCIDRNGGH--898
```

```

OY 1107 SPLQRYSEDPTVPLPSETGTYVAPLTCSPQPEYVNDPVRPQPPSPREGPLPAARAGAT 1166
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNYINVISLT 936
QY 1167 -LERAKTLPGRKNGVVKDVFAGGAVENPEYL 1197
DB 937 AISKLPMDSRYNQ-----SHSTAVDNPEYL 961

RESULT 14
OY 085468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC8C8CA0F8AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.2e-121;
Matches 345; Conservative 70; Mismatches 115; Indels 92; Gaps 15;

QY 578 GREADOCVACAHYKDPFCVACRPSGVKFDLSYMPIWKFDPDEGACQPCPINCTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPCTRGCKGP 57
QY 638 DDKGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRRQKIRKYTMRLLOETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGIGLGLYLR--HIVRKTLRLLQEREL 113
QY 697 VPLTPSGAMPNQAMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
DB 114 VPLTPSGEAPNQAHRLILKETEFKKVKVLSGAGFTVYKGLWIPGEKVTIPVAIKELR 173
QY 757 ENTSPKANKEILDEAYVMAGVSPVYVRLGLIGLCLTSTVQLTQLMFYGCLLDHVRENKGR 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCHLLGICLTSTVQLTQLMFYGCLLDYIREHKON 233
QY 817 LGSQDILLNWCQIAKMSYLEDLVRLVHRLDAARNVLKSPNHNKIPFGLA RLIDIDE 876
DB 234 IGSQYLLNWCQIAKGMNYSERHVRDLAARNVLKTPQHVKITDFGLAQQLGADKE 293
QY 877 YHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLE 936
DB 294 YHAGGKVPKIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLE 353
QY 937 KGERLPQPICTIDVYIMVWKCMIMDSECRPRFRELVSFSEFMRMARDPORFVIO-NEDLG 995
DB 354 KGERLPQPICTIDVYIMVWKCMIMDSECRPRFRELVSFSEFMRMARDPORFVIO-NEDLG 413
```

996	QY	PASPLDSTFYRSLLEDDMDGLVDABEYLVPQOGFFCPDPAPGAGGMVHRRHSSTRSG	1055
414	Db	LPSPTDSKFYRTLMEEDMEDIVDADLYLPHQOGF-----NSPST---454	
1056	QY	GGDLTLGLEFSBEEAPRSL-----APSEGAGSDVFDGLGMGAAGLQSLPHTHDPSPLQ	1110
455	Db	-----SKTPLLSSLSATSNNSATCNDRNG-----H-----481	
1111	QY	RYSEDPVPLPSETGDGVAPLTCSPQPEYVYNQDVRPQPSPREGPLPAAPACAT-LER	1169
482	Db	-----PVREDGFL-----PAPERYNQ-LMEKPKSTAMVQIQIYNYISLTAISK	523
1170	QY	AKTLSPGKNGVVKDVFAGGAVENPEYL	1197
524	Db	LPMDSRYON-----SHSTAVDNPEYL	544

RESULT 15

Q9WVF5	PRELIMINARY;	PRT;	655 AA.
AD	Q9WVF5		
AC	Q9WVF5;		
DT	01-NOV-1999 (TrEMBLrel. 12, Created)		
DT	01-NOV-1999 (TrEMBLrel. 12, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).		
DE	isoform 3).		
GN	EGFR.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
[1]			
SEQUENCE FROM N.A.			
RC	SPRAIN=C57BL/6J; TISSUE=LIVER;		
RC	Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.,		
RA	Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,		
RA	Maihle N.J.;		
RA	"Alternative Transcripts from the Human and Mouse EGFR Genes Encode		
RT	Carboxy-Terminal Truncated Receptors.";		
RT	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
SEQUENCE FROM N.A.			
RC	SPRAIN=C3H/101, 129/SVJ, AND 129/SVEVTCAC;		
RC	Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,		
RA	Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,		
RA	Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,		
RA	Maihle N.J.;		
RA	"Comparative genomic sequence analysis and isolation of human and		
RT	mouse alternative Egr transcripts encoding truncated receptor		
RT	isoforms.";		
RT	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
SEQUENCE FROM N.A.			
RC	SPRAIN=C57BL/6J; TISSUE=LIVER;		
RC	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Giessi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Harsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Wombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sesaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,		
RA	Hayashizaki Y.;		

"Functional annotation of a full-length mouse cDNA collection.";

RT Nature 409:685-690 (2001).

RL EMBL; AF124513; AAD44149.1; -

DR EMBL; AF124513; AAD44149.1; -

DR EMBL; AF275366; AAG28047.1; -

DR EMBL; AF275364; AAG28047.1; JOINED.

DR EMBL; AF275365; AAG28047.1; JOINED.

DR EMBL; AF275365; AAG28047.1; JOINED.

DR EMBL; AK004944; BAB23688.1; -

DR EMBL; AK004883; BAB23641.1; -

DR EMBL; AK004911; BAB23662.1; -

DR MGD; MGI:95294; Egrf.

DR InterPro; IPR000494; EGFR_L domain.

DR InterPro; IPR002174; Furin-like.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF01030; Recep_L domain; 2.

DR SMART; SM00261; FU; 3.

KW Receptor.

SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.9%; Score 1486.5; DB 11; Length 655;

Best Local Similarity 43.7%; Pred. No. 2.6e-103;

Matches 279; Conservative 98; Mismatches 239; Indels 23; Gaps 7;

QY 11 LLLALLPPGAA--STQVCTGTDMKRLPASPETHLMDLRHLRYGCGVQVQGNLELTLYLPTN 68

DB 14 LLLTALCAAGALBEKKVCGQTSNRLTQLGTFEDHFLSLQRYNNCEVLGNLEITVQRN 73

QY 69 ASLSFLQDIQEVQGYVLIHAHQVQVLPQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128

DB 74 YDLSFLKTIQEVAGYVLIALTNTVERIPLNQLIIRGNALYENTYALAILSN----- 124

QY 129 VTGASFGGLBELQLRSLTEILKGGVLIQRNPQLCYODTILWKDI----PHKNQLALTLI 184

DB 125 -YGTNRTGLRELPMRNLOEILIGAVFSNNPILCNMDTIQWRDIVQVFMNSNMDL--- 180

QY 185 DTNRSRACHPCSPWCKGSRGWSSDDCSLRTVCAGGCA-RCKGPLPTDCCHEOCAAG 243

DB 181 -QSHPSCKPKDSCPNFGSCWGGGEBNCQKLTIIQAQCSHRCGRSPSCDCHNCCAAG 239

QY 244 CTGPKHSIDCLAHFNHSGICELHCPALVOYIKANSKFIGITELRYTFGASCVTACPNY 303

DB 240 CTGPRESDCLVCKQFQDEATCKDTPPLMLYNPTTYQMDVNPBGKYSFGATCVKCKPRNY 299

QY 304 LSTDVGSCTILVCPLNHQEVTAEDGTORCEKCKPCARVCVGLGMEHLREVRVATSANIQE 363

DB 300 VVTDHGSVCRAACPDYDEV--EEDGIRKCKCDGFCRKNVCGNGIGIGBFKDTLSNATNIKH 358

QY 364 FAGCKKIFGSLAFLPESFGDPSASNTAPIQEQLOQVFLEETITGVLYTSAWPDSLPLDL 423

DB 359 FKYCTAISGLDHLHPVAFKGDSTRTPDLPRELEILKTVKEITGTFLLQAWFDNWDLLH 418

QY 424 VFQNLQVIRGRILIHNGAYSITLQGLGISMLGSLRSLRELGSGLAIHHNTHLCFVHTVPWD 483

DB 419 AFENLEIIRGTRTKHQGFQSLAVVGLNITSLSGLSLKEISDGDVIIISGNRLCYANTINWK 478

QY 484 QLFNPHQALLHTANPEDEECVEGLACHQLCARGHCWGPGPTQCNCISOFLRGSCVEE 543

DB 479 KLFGTNPQTKIMNNRAEKDCKAVNVCNPLCSSEGCWGPPEPRDCVSCQNVSRGRCVEK 538

QY 544 CRYVQLGLPREYNVARHCLPCHPECPQNGSVTCFGEADOCVACHYKDPFFCVACRCPSG 603

DB 539 CNTLEGEPRFVENSSEICQHPCELCQAMNITCTGRGPDNCIOCAHYIDGPHCVKTCPCAG 598

QY 604 VKPDLSPYPIWKFPFDEEGACQPCPINCTHSCVDLDDKGC 642

DB 599 IMGENTNL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636

Search completed: July 22, 2003, 09:00:50
Job time : 53.3575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 63.1875 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-4
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR: 73.*
2: PIR: **
3: PIR: 3.*
4: PIR: 4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6806	99.9	1255	1 A24571	protein-tyrosine k
2	5988	87.9	1260	1 TVRTNU	protein-tyrosine k
3	5984.5	87.9	1254	2 I48161	p-185 precursor -
4	3168	46.5	1210	1 GQHUE	epidermal growth f
5	3144	46.2	1210	2 A53183	epidermal growth f
6	3123.5	45.9	1223	1 TVCHLV	epidermal growth f
7	3003.5	44.1	1308	2 A47253	epidermal growth f
8	2701	39.7	1166	1 S06142	protein-tyrosine k
9	2431.5	35.7	1342	2 A36223	kinase-related tra
10	2346.5	34.4	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1652.5	24.3	1330	1 GQFHE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVEB	protein-tyrosine k
18	1536	22.5	644	2 A36325	epidermal growth f
19	1302	19.1	1323	2 B88257	protein let-23 (im
20	1302	19.1	1374	2 S70712	protein-tyrosine k
21	1214	17.8	1369	2 S70713	protein-tyrosine k
22	1177	17.3	1717	1 A45558	epidermal growth f
23	1155	17.0	527	2 A42032	epidermal growth f
24	997.5	14.6	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	735	10.8	1363	2 T43220	insulin-like growt
28	718	10.5	1382	1 INHUR	insulin receptor p
29	711	10.4	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB.

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A4188; B4188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saico, T.; J

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R;Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R;Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:G183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:BRB2; NCL: NEU; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;718-983/Domain: intracellular #status predicted <INT>
 F;726-734/Region: protein kinase homology <KIN>
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 99.98; Score 6806; DB 1; Length 1255;
 Best Local Similarity 99.88; Pred. No. 1.2e-270;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MELAAALCRWGLLALLP	60
DB	1	MELAAALCRWGLLALLP	60
QY	61	ELTYLPTNAGSLFLQDIQ	120
DB	61	ELTYLPTNAGSLFLQDIQ	120
QY	121	DPNNTPVTGASPGGLRE	180
DB	121	DPNNTPVTGASPGGLRE	180
QY	181	LTLLDNRSPACHPCSP	240
DB	181	LTLLDNRSPACHPCSP	240
QY	241	AACTGPKHSDCLACLH	300
DB	241	AACTGPKHSDCLACLH	300
QY	301	YNVLTSDVGSCTVCP	360
DB	301	YNVLTSDVGSCTVCP	360
QY	361	IQEFAGCKKIFGSLAF	420
DB	361	IQEFAGCKKIFGSLAF	420
QY	421	DLVSFQNLQVIRGILH	480
DB	421	DLVSFQNLQVIRGILH	480
QY	481	PWDLFRNPHQALLHT	540
DB	481	PWDLFRNPHQALLHT	540
QY	541	VEECRVLQGLPREYV	600
DB	541	VEECRVLQGLPREYV	600
QY	601	PSGVKPDLSYMPITW	660

RESULT 2

TVRTNU

protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, S.

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663; 'V', 665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680V/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.9%; Score 598; DB 1; Length 1260;
Best Local Similarity 87.7%; Pred. No. 2.8e-237;
Matches 1103; Conservative 50; Mismatches 102; Indels 2; Gaps 2;

```
QY 1 MELAAALCRWGLLALLPPGAASQVCTGDMKRLPASPEHDLMLRHLYQGCVVQGNL 60
DB 4 MELAAALCRWGLLALLPPGAGTQVCTGDMKRLPASPEHDLMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 64 ELTYVPAASLSFLQDIOEVQGYVLIHAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNR 123
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 179
DB 124 DPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQORNPOLCYQDMVLWKDIFRKNQNL 183
QY 180 ALTLIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 239
DB 184 APVDIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PYNLYSTDVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTS 359
DB 304 PYNLYSTEVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRGARITSD 363
QY 360 NIGEPAGCKIKFSLAPLPSFGDPSANTAPLOPEQLQVFTEILTEITGLYTSAMPDLSL 419
DB 364 NVQEFDFGCKIKFSLAPLPSFGDPSANTAPLOPEQLQVFTEILTEITGLYTSAMPDLSL 423
QY 420 PDLVSFQNLQVIRGRLIHNGAYSLTLQGLIGISWLGRLSRELGSGLALIHNNHLCFVHT 479
DB 424 RDLVSFQNLRIIRGRLIHNGAYSLTLQGLIGISWLGRLSRELGSGLALIHNNHLCFVHT 483
QY 480 VPWDQLFRNPHQALLHTANPEDE-CVGEGLACHQLCARGHCGPPTQCVNCSQFLRGQ 538
DB 484 VPWDQLFRNPHQALLHSGNRPEDLCVSSGLVCNLSLCAHCHGCGPPTQCVNCSHFLRGQ 543
QY 539 ECVEECRVGLQGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCVACAHYKDPDFCVA 598
DB 544 ECVEECRVWKGLPREYVSDKRCCLPCHPECPQNSSETCFGSEADQCAACAHYKDSSCVA 603
QY 599 RCPSGVKPDLISYMPINKFPDEEGACOPCPINCTHSCVDLDDKCPAEORASPLTSVAV 658
DB 604 RCPSGVKPDLISYMPINKFPDEEGICQPCINCTHSCVDLDERGCPAEORASPTFIATV 663
QY 659 VGLLVVVGWVGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKET 718
DB 664 EGVLVLLVLLVVVGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKET 723
QY 719 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 778
DB 724 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 783
QY 779 SPVVSRLIGICLTSTVOLVTQLMPYCCLLDHVENRGLSGDILLNKCQIAKMSYLE 838
DB 784 SPVVSRLIGICLTSTVOLVTQLMPYCCLLDHVENRGLSGDILLNKCQIAKMSYLE 843
QY 839 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIKWMALESILRR 898
DB 844 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIKWMALESILRR 903
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVRC 958
```

```
DB 904 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVRC 963
QY 959 WMIDSCRRFRRELVSERMRDPORFVVIQONEDLGPASPLDSTFYRSLLEDDDDMDLV 1018
DB 964 WMIDSCRRFRRELVSERMRDPORFVVIQONEDLGPASPLDSTFYRSLLEDDDDMDLV 1023
QY 1019 DAEYLVPOQGFPCPDPAFGAGVMHRRSSSTSGGGDLTLGLPESEEAAPRSLAPS 1078
DB 1024 DAEYLVPOQGFPCPDPAFGAGVMHRRSSSTSGGGDLTLGLPESEEAAPRSLAPS 1083
QY 1079 EGAGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQE 1138
DB 1084 EGAGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQE 1143
QY 1139 YNQSDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1198
DB 1144 YNQSDVPRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLT 1203
QY 1199 POGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGIDVFP 1255
DB 1204 PREGTASPHPSAFSPAFDNLVYWDQDPPERGAAPPSTFKGTPTAENPEYLGIDVFP 1260
RESULT 3
148161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa,
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:g747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif
Query Match 87.9%; Score 598.5; DB 2; Length 1254;
Best Local Similarity 87.6%; Pred. No. 3.8e-237;
Matches 1099; Conservative 58; Mismatches 97; Indels 1; Gaps 1;
QY 1 MELAAALCRWGLLALLPPGAASQVCTGDMKRLPASPEHDLMLRHLYQGCVVQGNL 60
DB 1 MELAAALCRWGLLALLPPGAGTQVCTGDMKRLPASPEHDLMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNR 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 180
QY 181 LTLIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 240
DB 181 PVDIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 301 PYNLYSTDVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTS 360
DB 301 PYNLYSTEVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRGARITSD 360
```


603 NNTL-VVKYADAGHVCHLCPNCTYCTCGPLGEGCTNGPKIP--SIATGMVGALLLLV 659
 665 VVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETELRKVK 724
 660 VALGIG--LFWERRHHVRKTRLRLQLQRELVEPLTPSGEAPNQALLRLKETEFKKIK 716
 725 VLGSAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELLDYAYWAGVGSPPVSR 784
 717 VLGSAGFTVYKGLWIPGEKVKIPVAIKRELREATSPKANKELLDYAYWASVDNPHVCR 776
 785 LIGICLTSTVOLTPQLMPYGCILLDHVRENRGLSGODLLNWCQIAKGMSYLEDVRLVHR 844
 777 LIGICLTSTVOLTPQLMPYGCILLDYVREHKDNIQSYLLNWCQIAKGMYLEDRLVHR 836
 845 DLAAARNVLKSPNHNKITDFGLARLLDIDETEHADGKGKVPKIKMALESILRRFTHQSD 904
 837 DLAAARNVLKTPQHVKITDFGLAKLLGAEKEYHAESGGKVPKIKMALESILHRIYTHQSD 896
 905 VWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICTIDYVMIMVKCWMIDSE 964
 897 VWSYGVTVWELMTFGSKPYDGIIPASBISISILEKGERLPOPPICTIDYVMIMVKCWMIDAD 956
 965 CRPRERELYSERSMARDPQRFVVIQ-NEDLGPASPLDSTFFYBSILLEDDDMGLVDAAEY 1023
 957 SRPKFRELITFSKWARDPQRYLYLVIQDGRMHLPSTDSNFYALMDEEDMDVDVDAEY 1016
 1024 LVPQQGFPCDPAPGAGGMVHHRHSSTRSGGGDLTLGLEPSEEBEAPRSPAPSEGAGS 1083
 1017 LI PQGGFF-----SSPSTRTPLLSLSATS 1042
 1084 DVPDGLDGMGAAGLQSLPHTDPSPLORYSEDPVPLPSET--DGVVAPUTCSQPPEYVN 1141
 1043 N--NSTVACIDIRNGLQSCPIKEDSFQRYSSDPTGALTEDSIDDTFL-----PVPEYIN 1094
 1142 QPQVRRQPPSPRGPLPAARPGATILERAKTLSPGKGVVVKVFAFGGAIVENPEYL-TPQ 1200
 1095 Q-SVPKRPAGSQVNPVYHQPLNP-----APSRDPHYQD--PHSTAVGNPEYLNVTQ 1143
 1201 GGAAPQPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPPSTFKGTPTAE 1244
 1144 -----PTCVNSTFDPAHWAQKSHQISLDNDPDYQDFFPKAKNGIIFKGS-TAE 1193
 1245 NPEYL 1249
 1194 NAEYL 1198
 RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C:Accession: A53183; A43818; S24942; A28943; 845325; I49643
 R:Luettecke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkinson Dev. 8, 399-413, 1994
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF r
 A:Reference number: A53183; MUID:94170986; PMID:8125255
 A:Accession: A53183
 A:Molecule type: mRNA
 A:Residues: 1-1210 <L0UE>
 A:Cross-references: GB:U03425
 R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand bi
 A:Reference number: A43818; MUID:91232866; PMID:2030916
 A:Accession: A43818
 A:Molecule type: mRNA
 A:Residues: 1-714 <AVI>
 A:Cross-references: GB:X59698
 R:Eisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A:Reference number: S24942
 A:Accession: S24942

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A:Reference number: 149643; MUID:93126380; PMID:7678348
A:Accession: 149643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 122-132 <RES>
A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 46.2%; Score 3144; DB 2; Length 1210;
Best Local Similarity 49.8%; Pred. No. 2.8e-121;
Matches 633; Conservative 170; Mismatches 359; Indels 110; Gaps 23;

Qy 11 LLIALLPGAA--STQVCTGTDMLRLPASPETHLMLRLHYLQCCQVQGNLELYLPTN 68
Db 14 LLTALCAAGGALBKKVCOGTSNRLTQGTGFEDHFLSLQRMVNNCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLQDIOEVQGVYLIHNOVROVPLRLRIVRGTLQFEDNVALAVLQNGDPLNNTTP 128
Db 74 YDLSFLTKIOEVAGYVLIANTVERIPLENLQIIRGNALYENTVALAILSN----- 124
Qy 129 VTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILMKDI----FHKNNQLALTLI 184
Db 125 -YGNRTGLRELPMNLQELILGAVRFNNPILCNMDTIQWRDIVQNVFNSNMDL--- 180
Qy 185 DTVRSRACHPCSPMKSCRWGESSEDCSLTRTVAGGCA-RCKGPLPTDCHEQCAAG 243
Db 181 -QSHPPSCPKCDPSCPNPSCWGGGEENCQKLTIKIICACQCSHRGRSPSDCHNQCAAG 239
Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRSDCLVCQKQFQDEATCKTCTPEMLYNPTTYQMDVNPGEKYSFGATCKVKCPNY 299
Qy 304 LSTDVSGCTLVCPILHNOQVTAEDQTCRCEKSPCARVCYGLGMEHLREYRAVTSANIQE 363
Db 300 VVTDHGSCVRACGPDYEV-EEDGIRKCKKCDGCRKVCNGVIGIGEPKDTLSINATIKH 358
Qy 364 FAGCKKIFGSLAFPEFSDGDPASNTAPLOEQLOVETLEETGYIYISAWPDSLPLDS 423
Db 359 FKYCTAISGDLHLPLFAFGKDSFTRTPPLDPRELEILTKVETGFLLIQAWPDNWTDLH 418

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A00643

R:Iax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A:Reference number: A27720; MUID:88261272; PMID:3260329

A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
R;Cross-references: GB:M20386
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NIL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-654/Domain: extracellular #status predicted <EXT>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.9%; Score 3123.5; DB 1; Length 1223;
Best Local Similarity 48.7%; Pred. No. 2e-120;
Matches 632; Conservative 175; Mismatches 345; Indels 145; Gaps 25;
QY 8 RWGLLLALLPGAA-----STQVCTGTDMLRLPASPETHDMLRHLHYGQCQVVOGNLE 61
DB 13 RGAALVLLLLGVALCSAVEBEKVCQGTNNKLTQLGHVEDHFTSLQRMYNCEVLSNLE 72
QY 62 LTYLPTNASLFLQDIOEGVYVLIHNOVRQVPLORLRIVRGTLQPLDNDVALAVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGVLIAMVDVIPLENLQIIRGNVLDNSFALAVLSNYH 132
QY 122 PLANNTPVTGASPGGLRELQRLSLEILKGVLIQRPOLCYODTILWKDIFHKNQAL 181
DB 133 -MNTQT-----GURELPMKRLSELINGVKISNNPKLNCMDTVLWNDIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKGRSGWSESSDCSLRTVTCAGGCA-RCKGPLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNCQTLTKVICAQCSGRGKGVPSDCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSHGICELHCPALYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKFRDDATCKDTCPLVLYNPPTYQMDVNPBGKYSFGATCVREC 302
QY 300 PYNLSTDVSGCTILVCPHNLQEVTAEDGTORCEKCKSPCARVCYGLGMEHLREVRVTS 359
DB 303 PHNVVVDHSGVRSNCNTDYEV-ENGVRKCKCDGLSKVCNGVIGIGELKGLSINAT 361
QY 360 NIOEFACCKIPGSLAPLPSFGDPSANTAPIQEOQLVFTELEITGYLTYISAMPDSL 419
DB 362 NIDSFNCKTNGDVSLPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQAWPDNA 421
QY 420 PDLVFNQLOVIRGRIHLNGAYSITLQGLISGLRLSRELGLALIHNTHLCEVHT 479
DB 422 TDLYAFENLEIRGTQKHQGYSLAVNLKIQSLGRLSKLEISDGDIAIMKNKLCYADT 481
QY 480 VPWDQLFRNPQALLHTANPEPECVGEGIACHQLCARGHCWGPQTQVCNCSQFLRGQE 539
DB 482 MNWRSUFATOSQTKIIQIENKNDCTADRHVCDPLCSDCVCGWGPFGPFCFFRQKE 541
QY 540 CVEECRVQLPREYNARHCLPCHPECQPNQ---SVTCFGEADQCACAHYKDPFPC 596

DB 542 CVKQCNIQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKCAHFDIPHC 601
QY 597 VARCPGVKPDLSYMPIMWKPDBEGACQPCPINCSTHSCVDLDDKGPAPAEORASPLTSIVS 656
DB 602 VRACPAVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTFSIAA 657
QY 657 AVV-GILLVVVGVVFGIILIKRRQQKIRKYTRRRLLQETELVEPLTPSGAMPNQAMRIL 715
DB 658 GVVGGLLCLVVGIGIGLYLRRR-HIVKRTLRLQLERELVEPLTPSGEAPNQAHLRIL 716
QY 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYVMA 775
DB 717 KETEFKVKVLGSGAGFTVYKGLWIPEGEKVKIPVAIKELREATSPKANKELIDEAYVMA 776
QY 776 GVGSPVSVRLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGDQLLNMCMQIAGMSY 835
DB 777 SVDNPHVCHLLGICLTSTVQLITQMLPYGCLLDYIREHKDNTGSGVLLNWCQIAGMNY 836
QY 836 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALLESIL 895
DB 837 LEERRLVHRDLAARNVLVKTPOHVKITDFGLAKLGADEKEYHAEGKVPPIKWMALLESIL 896
QY 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIM 955
DB 897 HRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPPICTIDVYIM 956
QY 956 VKCWMIDSCRRFRRLVSEFSRMDPQRFVVIQ-NEDLGASPLDSTFYRLSLDDDDM 1014
DB 957 VKCWMIDASRPKRELIAFBNAMDPPYLIQODERHLPSPPTDSKYRTLMEEEDM 1016
QY 1015 GDLVDAEYLVFQQGFCFDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
DB 1017 EDIVDAEYLVFQGGFF-----NSPST-----SRTP 1042
QY 1075 L-----APSEGAGSDVFDGDLGMAAKGLQSLPTHTDPSPLQRYSEDPVTPLPSET--DGY 1127
DB 1043 LLSLSATSNNATNCID-----RNGQGHFVREDSEFQVSYSDPTGNFLEESIDDFG 1094
QY 1128 VAPLTCSPQEVYNQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPCKNGVVKDVF-- 1185
DB 1095 L-----PAPEVYQ--LMPKKPS-----TAMVQNIYNNISLT 1125
QY 1186 -----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQ----- 1225
DB 1126 AISKLPMSDSRYQNSHSTAVDNPEYL-----NTNQSPKLVTFESSFYWQSNGHQIN 1177
QY 1226 -DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
DB 1178 LDNPDYQQDFLNPETKPNGLLKVPAAENPEYLRVAAP 1214
RESULT 7
A47253
Epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plowman, G.D.; Culousecou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal s
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLQ>
A;Cross-references: GB:L07866; NID:g337359; PID:AA859446.1; PID:g337360
A;Note: sequence extracted from NCBI backbone (NCBIP:126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

Db 123 YQK-NPSSP--DYQVGLKQLQLNLTEILSGGVKSHNPLLCNVETINWWDIVDKTSNP 179
Qy 180 ALTLDNRGRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGC-ARCKGPLTDCCHE 238
Db 180 TMLNLI PHAFERQCKDHGCVNGSWAPGPGCHQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
Qy 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVA 298
Db 240 HCAGGCTGPRATDCLACRDNDDGCTCKDTPPPKIIDYVSHQVVDNPNIRKTYTGAACVKE 299
Qy 299 CPVNYLSTDVSGCTLVCLPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVATS 358
Db 300 CPSNYVYTE-GACVRSACSAGMLEVD-ENGKRSKPCDGVCPKVCDDGIGISLNTIAVNS 357
Qy 359 ANIQERAGCKKIFGSLAFLESFDDPASNTAPLQBPOLQVFTLEBITGYLYISAWPDS 418
Db 358 TNIRSFNCTKINGDIILNRNSPEGDPHYKIGTMDPEHLNLTIVKESITGYLVIMWVPE 417
Qy 419 LPDLSVFQNLQVIRGRILHNGAYS-LTLOGLIGISWGLRLSRLRELGLSLALHNTLHLCFV 477
Db 418 MTSLSVFQNLLEIRGRTTFRSGFSFVVQVRHLQWLGLRLSRLKEVSAGNVILKNTLQLRYA 477
Qy 478 HTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQCHGCPGPTOCVNCQSLRFG 537
Db 478 NTINWRLLFRSEDOSIEYDART-----ENQTCNECEDGCM-POPTMCVSLCHVDGR 529
Qy 538 QECVECRVLQGLPREYVYNARHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDDPPFCV 597
Db 530 GRCVASCNLLQGEPREAQVGRVCQCHQECVLTQDSLTCTGPGPANCSSAHPQDGPQCI 599
Qy 598 ARCPGKVPDLNYPWKFPDEGACQPCPINCTHSCVDLDDKGCPAEQASPLTSIVSA 657
Db 590 PRCPHGILGSDTL-IWKYADKMGQCQCHQNCCTQCGSGPLSGCRGD-IVSHSSLAVAL 647
Qy 658 VVGILLVVLGVVFGIILKERQKIRKYTWRLLOETELVEPLTPSGAMPNQAMRLIKE 717
Db 648 VSGLLIIVALLIVLLRRRIK-RKRTIRCLLQEKLEVEPLTPSQAPNQAFRLIKE 706
Qy 718 TELRKVKLGSGAGFYVYGIWIPDGENVKIPVAIKVLRNTPSPKANKETILDAYVMAGV 777
Db 707 TEFKDKRVLGSAGFYVYGLMNPDDGENIRIPVAIKVLRNTPSPKQNVQVLDAYVMASV 766
Qy 778 GSPVSRLLGICLTSTVQLVTQMPYGCLLDHYRENGRLGSQLLNCWQIAKMSYLE 837
Db 767 DHPHVCRLGICLTSAVLQVTQMPYGCLLDYVRQHOERICGOWLLNCWQIAKMSYLE 826
Qy 838 DVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILR 897
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLJTADKEYQADGGKVPKIMMALESILQW 886
Qy 898 RFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVK 957
Db 887 TYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASVLENGERLPQPPICITIEVYIMLK 946
Qy 958 CWMIDSECRPRRELVESEFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDGL 1017
Db 947 CWMIDSSRRPRRELVEGEFQMARDFRYLVIQG---NLPSLSDRLFRSLSDSD--DV 1001
Qy 1018 VDAEYVLVPOGFCFPPAPGAGGVVHRRSSSTRSGGSDLTILGLEPSEEAAPRPLAP 1077
Db 1002 VDAEYLLPYKRI-----NRQGS-----EPCIP 1024
Qy 1078 SEGAGSDVFDGLGMGAAGKQLSLPTHDPSPLORYSEDPTV-PLPSETDGVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSITLRNISTDPTQNALKDLGDH----- 1055
Qy 1137 PEYVQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGGVYKD 1183
Db 1056 -EYVNOPGSETSRSLSDIYNPNVEDLTDGWPVSLSSQEAETFSRPEYLNTQNSL--- 1111
Qy 1184 VPAFGGAVENPEYLTQGGAAPOBHPPPAPSPAFDNLVYWDQDPPEGAPPSTFKGTPTA 1243
Db 1112 PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPAA 1149

Qy 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223

kinase-related transforming protein (erbb3) (EC 2.7.1.-) precursor - human

C;Species: Homo sapiens (man)

C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C;Accession: A36223; I59164

R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A;Reference number: A36223; MUID:90083234; PMID:2687875

A;Accession: A36223

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1342 <KRA>

A;Cross-references: GB:M29366

R;PLOWMAN, G.D.; WHITNEY, G.S.; NEUBAUER, M.G.; GREEN, J.M.; McDONALD, V.L.; TODARO, G.J.

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A;Title: Molecular cloning and expression of another epidermal growth factor receptor-rel

A;Reference number: I59164; MUID:90311312; PMID:2164210

A;Accession: I59164

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A;Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C;Genetics:

A;Gene: GDB:ERBB3; HER3

A;Cross-references: GDB:119880; OMIM:190151

A;Map position: 12q13-12q13

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C;Keywords: ATP; phosphotransferase

F;707-972/Domain: protein kinase homology <KIN>

F;715-723/Region: protein kinase ATP-binding motif

Query Match 35.7%; Score 2431.5; DB 2; Length 1342;
Best Local Similarity 40.7%; Pred. No. 3.7e-92;
Matches 533; Conservative 191; Mismatches 458; Indels 129; Gaps 32;

Qy 10 GLLALLPPGAA--STQVCTGTDMKLRPLASPETHLMLRLHYLQGVQGVQGNLELYLPT 67

Db 11 GLLFSLARGSEVGNQAVCPGTLNGLSVTGDENQYQTLVLYRCEVVMGNLEIVLTGH 70

Qy 68 NASLSPLQDIQEVQGVYLAHQVQVPLQRLRIVRGTOIFEDNYALAVLDNGDPLNNTT 127

Db 71 NADLSFLQWIREVTGYVLVAMNEFTLPLNLRVVRGTQVYDGKFAIFVM-----LYNNT 125

Qy 128 PVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILMKDIFPKKNQLALTLDTN 187

Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNDKLCHMDITDWRDIVDRD---AEIVVKD 178

Qy 188 RSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGC-ARCKGPLTDCHEOCAAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RCWPGSDEDCTLTCTTICAPQCNGHCFCGPNPQCCHDECAGCSG 237

Qy 247 PKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVACTACPYNYLST 306

Db 238 PQDTCDFACRHFNDSGACVPCPQPLVYNKLTFLQLEPNPHTTKYQGVGVASCAPHNFV-V 296

Qy 307 DVGSCTLVCLPLHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVATSANTOEFAG 366

Db 297 DOTSCVRACPPDKMEVD-KNGLMKMEPCGGLCFKACEGTGSG--SRFTVDSNIDGFVN 353

Qy 367 CKKIFGSLAFLESFDDPASNTAPLQBPOLQVFTLEBITGYLYISAWPDSLPDLSVFO 426

Db 354 CTYKILGNLDFLTGLNGDPWHKIPALDPEKLVNFRVITREITGYLNIQSPPHMNFVSFS 413

Qy 427 NLQVIRGRILHNGAYS-LTLOGLIGISWGLRLSRLRELGLSLALHNTLHLCFVHTVPWQOL 485

Db 414 NLTTIGRSYLRNGFSLIMKLNVTSLGFRSLKEISAGRIYISANRQLCYHHSLNWKV 473
QY 486 FRNPHQALLHTA-NRPDECEVGEGLACHOLCARGHGWPGPTQCVCNCSQFLRGQECVEEC 544
Db 474 LRGTTERELDKNRPDRDVAEKVCDPLCSSGGCGWPGGQCLSCNRYSGRGVCVTHC 533
QY 545 RVLOGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFCVARCPGV 604
Db 534 NFLNGEPREFAHEAECSCHPECPMEGTATCNGSGSDTCAQCAHPRDGPCHVCSSCPGV 593
QY 605 KPDLSYMPINKFPDEBACQPCPNCTHSCVDLDDKCCPAEORA-----SPLTSIVSAVG 660
Db 594 LG--AKGPIKYPDVQNECRCHENTCGCKGPELQCLGTLVLKIKLTHLWALVIAG 651
QY 661 ILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTELVEPLTPSGAMPNOAQMRILKETE 719
Db 652 --LVVIFMMLGGTFLYWRGRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIFKETE 708
QY 720 LRKVKVLGGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVS 779
Db 709 LRKLVKLVGGVFTVHKGVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLDH 768
QY 780 PYVSRLIGICLTSTVQLVTQMLPVGCLLDHVRNRRGLSQDILLNWCMTAKGMSYLEDV 839
Db 769 AHVRLGLCLPGSSLOLVTOYPLGLSLLDHVRQHRGALGPQLLLNMGVQIAKGWYILEEH 828
QY 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMWALESLIRRF 899
Db 829 GWHRNLAARNVLKSPQVQADFGVADLLPPDDKQLLYSEAKTPIKMMWALESIHFGKY 888
QY 900 THQSDVMSYGVYVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPCTIDVVMIMVKW 959
Db 889 THQSDVMSYGVYVWELMTFGAEPYAGRLAEPVDPDLLEKGERLAQPOICTIDVVMIMVKW 948
QY 960 MIDSECRPRRELVSERFARQDQRFVITQNEDLGPA---SPLDSTFYRSLLEDDMDG 1016
Db 949 MIDENIRPTEKELANEFTRWARDPPRYLVIKRES-GFGIAPGPEPHGLTKLEEEVELEP 1007
QY 1017 LVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLEP-SEEEAPRSL 1075
Db 1008 ELDDLDDLEAED-----NLATTLGSLSLPVTGNRRPFGSOLL 1048
QY 1076 APSEGAGSDVFDGLGMAAGLQSLPHTD-PSPLQRYSEDPTVPLP-----SETDGYV 1128
Db 1049 SPSSGY-MPMNOGNLGSQESAVSGSSERCPVSLH-----PMRPGCLASSESGHV 1101
QY 1129 A-----PLTCSPOPE-----YVNPQDVRPOPSPREP-----L 1157
Db 1102 TGSEAELOEKVSMCRSRSRSPRPGDSAYHSQRHSLLTPTVPLSPGLEEEDVNGYV 1161
QY 1158 PAARPAGATLERAKTLP-GKNGV-----KOVFAFGGAVENPEYLTPOGGAPOPHPP 1210
Db 1162 PDTHLKTSPSREGTLLSVGLUSSVLGTEEDED-----EYEYNNRRRRHSP-PHPP 1212
QY 1211 PAFSPAFDNLVYMD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
Db 1213 RPSLEELGVEYMDVGSLSASLGTSQCLPHVPVIMPTAGTTPDEYEM 1263

RESULT 10
JC4387

epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID: 96096535; PMID: 8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390

A;Experimental source: liver

A;Note: The authors translated the codon AAC for residue 369 as Thr and Grr for residue 1;
C;Comment: This protein is a functional heregulin receptor that transduces signals to the

C;Genetics:

A;Gene: ErbB3

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog;
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>

F;705-970/Domain: protein kinase homolog <KIN>

F;713-721/Region: protein kinase ATP-binding motif

F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cc

Query Match 34.4%; Score 2346.5; DB 2; Length 1339;

Best Local Similarity 40.8%; Pred. No. 1.1e-88;

Matches 523; Conservative 171; Mismatches 434; Indels 155; Gaps 34;

QY 3 LAALCRWGLLLALLPPGAA---STVCTGTDMKRLPASPETHLDMRLHLYQCGQVVOGN 59

Db 7 LQVLC-----FLSLARGSEMGNSQAVCPGTNLGLSVTGDADNQYQTLYKLYKECEVMGN 62

QY 60 LEITYLPTNASLFLQDIOEVQYVLIHNRQVPLQRLRIVRGTLQDFEDNVALAVLDN 119

Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNFSVLPLNLRVVRGTQVYDVGKFAIFVM-- 120

QY 120 GDPNNTTPTVGASPGGLRELQRLSLEILKGGVLTORNPOLCYQDTILWKDIFHKNNQL 179

Db 121 ---LNTNT---NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCMDTIDWRDVRVR--- 170

QY 180 ALTLIDNRSRACHPCSPMKSGRSGESSEDCQSLTRTVACAGGC-ARCKGPLPTDCCHE 238

Db 171 GAEIVVKNANGANCPCHEVCKG-RCWGPGPDCCQLITKTCAPQCNCRGCFGNPNCCHD 229

QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298

Db 230 ECAGGCGPDQDCFAACRRFNDSGACVPCPEPLVYNKLTFLQEPNPHTKYQYGGVCVAS 289

QY 299 CPNYLSTDVGSCTLYCPLHNEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTS 358

Db 290 CPNHFV-VDTQFCVRACPPDKMEVD-KHGLKMPCECGGLCPKACEGTSGS--SRYTQVDS 345

QY 359 ANIQEPAGCKIKFGSLAFLEPESFDGDPASNTAPLOEQLOVFEETLEITGLYVISAAPDS 418

Db 346 SNIDGFVNCTKILGNLDLITGLNVDPWKIPALDEKLNVRFTVREITGLYNIQSWPPH 405

QY 419 LPDLSVFONLQVIRGIRILHNGAYS-LTLOGLGISWGLRSLRGLSGLALIHHTHLCFV 477

Db 406 MNFVSFVSNLTITGSRSLYNRGFSLLIMKNLVNVTSLGFRSLKEISAGRVVISANQQLCVH 465

QY 478 HTVPMDQLFRNPHQALLHTA-NRPDECEVGEGLACHOLCARGHGWPGPTQCVCNCSQFLR 536

Db 466 HSLNLTLLRGLRPSSEELDIKYDRPLGECLEAGKVCPLCSCGGCGWPAFPQCLSCSNYSR 525

QY 537 GQCEVBEERVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPDFC 596

Db 526 EGVVTHCNFLOGEPREFVHEAQCFCHPECLPMEGTSTYNGSGSDACARCAHFRDGPCH 585

QY 597 VARCPGVKPDLSYMPIMKFPDEBACQPCPNCTHSC--VDLDDKGCFAEQRASPLTSI 654

Db 586 VNSCPHGILG--AKGPIKYVPOAQNECRCHENTCGCGPELQDCLGQAEVLMSKPHLV 643

QY 655 VSAVGVILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTELVEPLTPSGAMPNOAQMR 713

Db 644 IAVTVG--LAVILMILGSGFLYWRGRIQNKRAMRRYLERGESIEPLDPS-EKANKVLAR 700

QY 714 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYV 773

Db 701 IFKETELRKVLGSGVFTVHKGIWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLA 760

QY 774 MAGVGPVYSRLIGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSQDILLNWCMTAKGM 833

Db 761 VGSIDHAHIVRLILGLCPGSSQLVQYPLGLSLDDHVHQHRETLGQQLLNWGVQIAKGM 820
QY 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKMALES 893
Db 821 YLEESHMVHRDLAARNVLMKSPSQVQVADFGVADLLPDDKQLLHSEAKTPIKMALES 880
QY 894 ILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYM 953
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLAEIPDLLEKGERLAQPOICTIDVYM 940
QY 954 IMVKCMWIDSECRPRELVSERFWARDPQRFVWQNEQDLGASPLDSTFYRSILLEDD 1013
Db 941 VMVKCMWIDENIPTKELANETFRWARDPRVILVIKGRAS-GRGTP--PAEFSVITKE 997
QY 1014 MGDVDAEYLVPOQGFPCPDPAAGAGGMVHRRHSSTRSGGDLTLGLEPSEE----- 1068
Db 998 L-----QAELEPEL-----DLDLDEAESEGLATS 1023
QY 1069 -----EAPRSLAPSEG-----AGSDVDFDGLGMAKGLQSLPTH 1105
Db 1024 LGSALSPLTGTLTRGQSQSLSPSSGYMPMNQSSILGEACLDGSAVLGREGFRPISLH- 1082
QY 1106 PSLQRYSEDPVPLPSETDGYV-----APL-----TC-----SPOPE---YVNPQDV 1145
Db 1083 PIPRGR-----PASESEGHVTGSEAELOEKVUSVCRSRSRSPRPGDSAYHSQRHS 1135
QY 1146 RPQPPSPREGP-----LPAARPAAGATLERAKTLSP-GKNGV-----KDVFAF 1187
Db 1136 LLTPVTLSPGLEEDGNGVMPDTHLRCASSRRECTGLSSVGLSGTTEDEDED----- 1191
QY 1188 GGAVENPEYLTPOGGAAPOHPHP 1210
Db 1192 -----EEYENMRKRGSP-PRPP 1209

RESULT 11
TVFVLU
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.2%; Pred. No. 2.6e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 578 GPEADQCVACHYKDPFCVARGCPGVKPLDSYMPKFPDPDEGACPCPINCTHSCVDL 637
Db 60 GP--DHCMMKAHFDGPHCVKACPAVGLGENDTL-VWKYADANAVCOLCHPNCTRCKGP 116
QY 638 DKGCPAERASPLTSTVSNAV-GILLVVLGVVFGILIKRROOKIRKYMRLLQETEL 696

Db 117 GLEGCP---NGSKTFSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKETLRLLQEREL 172
QY 697 VEPLTPSGAMPNQAOQMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLR 756
Db 173 VEPLTPSGEAPNQAHRLILKETEFKKVKVVLGSGAFGVYKGLWIPGEKVKIPVAIKELR 232
QY 757 ENTSKANKKEIIDEAIVMAGVSPVYSRLILGICLTSTVOLVTQLMPEYGCGLLDHVRENRR 816
Db 233 EATSPKANKKEIIDEAIVWASVDNPHVCRLLGICLTSTVOLITQLMPEYGCGLLDYIREHKON 292
QY 817 LGSQDLNWCWMOIAKMSYLEDLVHRDLAARNVLKSPNHNKITDFGLARLLDDIDE 876
Db 293 IGSQVLLNWCVOIAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLAKLLGADEKE 352
QY 877 YHADGSKVPIKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGSKVPIKMALESILHRYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
QY 937 KGERLPPOPICTIDVYIMVKCMWIDSECRPRELVSERFWARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPPOPICTIDVYIMVKCMWIDADSRKPRELJAEFSKWARDPPRYLVIQSDERMH 472
QY 996 PASPLDSTFYRSILLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGGMVHRRHSSTRSG 1055
Db 473 LPSPTDSKPYRTLMEEEDMEDIVDADEYLVPHQGF-----NSPST--- 513
QY 1056 GGDLTILGLEPSEEEAPRSP-----APSEGAGSDVDFDGLGMAKGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLSSLSATSNSATNCID-----RNGQHPVREDSFVQ 550
QY 1111 RYSEDDTVPLPSET--DGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLE 1168
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKXPS----- 585
QY 1169 RAKTILSPGKNGVVKQVF-----AFGGAVENPEYLTPOGGAAPOHPHPPAF 1213
Db 586 -----TAMVQNIYNISLTAISKLPMDSRYQNSHSTAVDNPEVL-----NTNOSPLA 633
QY 1214 SPAFONLYYWDQ-----DPPE-----RCAPPSTFGTPTAENPEYLGLDVP 1254
Db 634 KTVFESSPYIQSGNHQINLNDPVDQDFLPNETRENGLLKVPAAENPEYLRVAAP 689

RESULT 12
TVFVLU
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of c
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28, 'W', 30-139, 'P', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.9e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 587 CAHYKDPFPCVACPSGVKPDLSVMPYTKWPKDEGACQPCPINCSTHSCVLDKDGKPAEQ 646
DB 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVQCLCHPNCRTGCKGFLGEGCP-- 58
QY 647 RASPLTISVAVV-GILLVVLVGVVGLIKRROOKIRKVTMRLLQETELVPLTPSGA 705
DB 59 NGSKTPTSIAGVGGCLLVVGLGIGLYLRR-HIVRKTLRLRLQERELVPLTPSGE 117
QY 706 MPNOAQRILKETELRKVKVYLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANK 765
DB 118 APNOAHLRIKETEFKVKVYLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANK 177
QY 766 EILDEAVVMAGVSPYVSRILGICLTSTVQLVQIMPYGCLLDHVRNRLGSGDILLNW 825
DB 178 EILDEAVVMASVDPNPHVCRLLGICLTSTVQLVQIMPYGCLLDYIREHKDNIGSQVLLNW 237
QY 826 CMQIAKMSYLEDLVRLVHRLAARNVLKSPNVKITDFGLARLLDIDETEXHADGKVP 885
DB 238 CVQIAKGMVLEERLVHRLAARNVLKTPQHVKITDFGLAKLGLADEKEYHAEGKVP 297
QY 886 IKWALESIILRRFTHQSVDVMSYGVTVWELMTFCAKPYDGIPIAREIPDLLEKGERLPQPP 945
DB 298 IKWALESIILRIYTHQSVDVMSYGVTVWELMTFSGKEYDGIPIASEISSVLEKGERLPQPP 357
QY 946 ICTIDVYIMVKWIMIDSECRPRELVSEFSRMARDPQFVIO-NEDLGASPLDSTF 1004
DB 358 ICTIDVYIMVKWIMIDSECRPRELVSEFSRMARDPQFVIO-NEDLGASPLDSTF 1004
QY 1005 YRSILLEDDMDGLVDAEYLVPOQFPCDPAPGAGGWHHRSSSTRSGGDLTLGLE 1064
DB 418 YRTIMEEDMEDVDAEYLVPOQF--NSPT----- 449
QY 1065 PSBEEAPRSP-----APSEGAGSDVFDGLGMGAAGLQSLTHPSPLOQYSEDPVP 1119
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPREDSFVORYSSDPTGN 495
QY 1120 LPSET--DGIVAPLTCSPQBYVNPQVRRPQSPREGPLPAARPAGATLERAKTILSPGK 1177
DB 496 FLEESIDDDGL-----PAPEYVNO--LMPKPTSTAM----- 524
QY 1178 NGVYKDVFAF-----GGAIVENPEYLTPOGGAAPQHPPPAFSPAF 1218
DB 525 --VQNIYNFISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNQSPKATVFE 574
QY 1219 NLYYWDQDPPERGAPPSTFGTPTAENPEY 1248
DB 575 SSPYWIQSGNHQ-----INLDNPDY 594
RESULT 13
GQFPE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlesinger, J.; Shilo, B. Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <Liv>
R;Wadsworth, S.C.; Vincent III, W.S.; Billoreau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021

A;Molecule type: DNA
A;Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK', <WAD>
A;Cross-references: EMBL:X02293; NID:G7922; PIDN:CAA26157.1; PID:G929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57F
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122,300,324,363,518,688,695,700/Binding site: carbohydrate (asn) (covalent) #status pre
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte
Query Match 24.3%; Score 1652.5; DB 1; Length 1330;
Best Local Similarity 29.9%; Pred. No. 2.2e-60;
Matches 413; Conservative 180; Mismatches 415; Indels 371; Gaps 39;
QY 80 VQGVYLIHNVQVRQVPLQRLIRIVRGTLF-----EDNYALAVLDNGDPLNNTTPVTGASP 134
DB 38 ITNYIVIGLDLIPCTLSYRLQIRGTLFSLSVEEKYALFV-----TY 81
QY 135 GGLRELQLRSLTEILKGVLIQINPOLCYODTILWKIDIFHNQNLALTIDNRSRACHP 194
DB 82 SKMYTLEIPDLADVLNGVGFHNHNLCHMTIOWSEIVSNGTAYNYDFDTAPERCEPK 141
QY 195 CSPMKGSKWCSESSDQSLRTVTCAGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDC 252
DB 142 CHESCTHG--CWEGPKNCQKFKSLTCSQACGRCYGPKECHFCFAGGCTGPTQKDC 200
QY 253 LACLHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLSIDVGSCT 312
DB 201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETPNPEGYAYGATCVKECP--CHLLRDNGACV 259
QY 313 LVCPLHNQEVTAADGTQRCCKSPCARVCYGLGMHLEHVRVAVTSANTQEFAGCKKIFG 372
DB 260 RSCPDQKMDKGE-----CVPCNGPCPKTCPCGTVLH-----AGNIDSFRNCTVIDG 306
QY 373 SLAFIPESPDG--DPASNTA-----PLOPEQLQVPELLEITGYLYISAWPDSLPLSV 424
DB 307 NRIIDQTFSGQDVVYATMGFPRIPLDPERREVFSTVKEITGYLNIETHPQFNLSY 366
QY 425 FQNLQVIRGRILHNGAY-SLTQLGLGISWGLRSLRSLRELGLALIHNTHLCFVHTVPWD 483
DB 367 FRNLETHGRQLMESFAALAIVKSSLSYLEMRNLKQISSGVVIOHNRDLCVVSIRWP 426
QY 484 QLFPRPHQALLHTANRPEDE----- 504
DB 427 AIQKEPEQKVWVNNENRADLCGKFLTILISVQHNIIMHIFAICREKWNHLLAGSVQRGLL 486
QY 505 ----- 504
DB 487 GSWHGSVPYLOELQFQWHLHRLRLWLYQVINSITQDKSNEHQLTDACYSFSPVTSUTIER 546
QY 505 -----VGEGLA-- 510
DB 547 ARYATQSAGLAMELEQITARSASMRHSKTLPAEGRQVPRVFLGVCA SARAGIAEPLAGR 606
QY 511 -----CHQICARGHCWGPGPTQVCNCSQFLRQECVCECRVLOGLPREVY---NARHCLP 562
DB 607 AVCRKCHPLCELCTNYGVHEQVCSKCTHYKREQCETEC-----PADHYTDBEQRECFCQ 660
QY 563 CHPEQCPQNSVTCFGEPAQDCVACAHYK-----DPPF-----CVARCPSGVK--PDL 608
DB 661 RHPEC---NG---CTGPGADDCCKRNFKLFDANETGYPYVNSTMTNCTSKCPLEMRHNY 714
QY 609 SYMPIWKFPDEEGACQPCPCINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVGLLVVVLG 668

Db 715 QYTAIGPY-----CAASPPRSSKITANLD-----VNMIFITGAVLVPTIC 755
QY 669 VFGI-LIKRQOKIRRYT--MRLLQETELVPLTPSGAMPNQAOAQRILKTELAKRVK 725
Db 756 ILCVVTYICQKQKAKKETVMTWALSGRESEPLRPSNIGANLCKLIVKDAELRGV 815
QY 726 LSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGSPPVSR 785
Db 816 LGMGAFGRVYKGVVPEGENVKIPVAIKELKSTGAESSEFEFLREAYIMASBEHVLLKL 875
QY 786 LGICLSTVOLVTLQMPYGLLDHVRNRLGSLDILANCMQIAKMSYLEDLVLRH 845
Db 876 LAVCMSQMLITQLPLGCLLDYVRNNDKIGSKALLNWNSTQIAKMSYLEEKRLVHRD 935
QY 846 LAARNVLVK---SPNVHVKITDFGLARLLDIDETEHADGGKVPKIMWALESIILRRFTHQ 902
Db 936 LAARNVLRLAGEDH---DFGLAKLLSSDSNEYKAAGKMPKWLALCIRNVFTSK 991
QY 903 SDVWSGVTVWELMTGAKDYDGPAREIPDLLEKEGERLPQPPICITIDVYIMVIMVKMID 962
Db 992 SDVWAFGVITWELTQGRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLD 1051
QY 963 SECRRPREFLVSFMRWARDPQRFVVIQNEGLG--PASPLDSTFYRSLLEDD---DWGDL 1017
Db 1052 AAMRPTFKQLTTFVAFARDPGRYLAIGDKFTRLEA-----YTSQDEKOLIRKLAPT 1104
QY 1018 VDAEYELVPOQGFPCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEERAP----- 1071
Db 1105 TDGSEAIKPDYDLQKALGPS-----HRTDCT-----DENPKLNRVC 1143
QY 1072 RSLPASEGAGSDVFDG---DLGMGAAGLQSLPHTDPSLPORYSEDTVPPLSETDGYV 1128
Db 1144 KPSNKNSSGDDERDSSAREVGVGNLR-----LDLPVDEDDYL 1182
QY 1129 APLTCSPOEVYVQPDVVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKVFAFG 1188
Db 1183 MP-TCQPGNNNNNMN-----NPNQNNMAVGAAGYM-----DLIGVP 1220
QY 1189 GAVENPEYL---TPQGAAPQPH-----PPPAFSP-AFONLYYWD 1224
Db 1221 VSDNPEYLLNAQTLGVGSEPIQTQIGIPVWGPGTMEVKVMPGSEPTSSDHEYND 1279

RESULT 14

S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C>Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.5e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADOCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAAGLVGENDTL-VMKYADANAVCOLCHPNCTGCKGP 57

QY 638 DKGCPAEPQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 696
Db 58 GLEGCP---NGSKTFESIAAGVVGGLLCLVVVGLIGLYLRRR-HIVKRTLRLLQEREL 113
QY 697 VPPLTPSGAMPNQAOAQRILKTELKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VPPLTPSGAMPNQAOAQRILKTEFVKRVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLR 173
QY 757 ENTSPKANKEILDEAYVMAGVGSPPVSRILGICLSTVOLVTLQMPYGLLDHVRNRRGR 816
Db 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLSTVOLVTLQMPYGLLDHVRNRRGR 233
QY 817 LGSQDLLNCKMOIAKMSYLEDLVLRHDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNVCVQIAKGMNLYLEERHVMHVRDLAARNVLVKTPOHVKITDFGLAKQLGADKE 293
QY 877 YHADGGKVPKIMWALESIILRRFTHQSDVWSGVTVWELMTGAKDYDGPAREIPDLLE 936
Db 294 YHAGGGKVPKIMWALESIILRRFTHQSDVWSGVTVWELMTGSKPYDGPASEISVLE 353
QY 937 KGERLPQPPICITIDVYIMVIMVKMIDSECRPFRELVSFMRWARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPPICITIDVYIMVIMVKMSDADSRPKFRELIAEFKWARDPPRYLVIOGDERMH 413
QY 996 PASPLDSTFYRSLLEDDMGDLVDAEYELVPOQGFPCPDPAAGAGVHHRHRSSTRSG 1055
Db 414 LPSPTDSKYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLTGLGLEPSEERAPSL-----APSEGAGSDVFDGDLGMGAAGLQSLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNSNATNCIDRNGG-----H----- 481
QY 1111 RYSEDPVPLPSETDGYVAPLTCSPQEVYVQPDVVRPQPPSPREGPLPAARPAAGAT-LER 1169
Db 482 -----PVREDFGL-----PAPEYVQ--LMPKFPSTAMVQVQNIYVLSLTAISK 523
QY 1170 AKTLSPGKNGVVKVFAFGGAVENPEYL 1197
Db 524 LPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus (1
C:Species: avian erythroblastosis virus
C>Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 578 GPEADOCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAAGLVGENDTL-VMKYADANAVCOLCHPNCTGCKGP 57
QY 638 DKGCPAEPQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKYTMRLLOETEL 696
Db 58 GLEGCP---NGSKTFESIAAGVVGGLLCLVVVGLIGLYLRRR-HIVKRTLRLLQEREL 113

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 33.0519 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703A-4

Perfect score: 6812

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6806	99.9	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5994	88.0	1237	1 ERB2_RAT	P06494 rattus norv
3	5984.5	87.9	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3166	46.5	1210	1 EGFR HUMAN	P00533 homo sapien
5	3145	46.2	1210	1 EGFR MOUSE	Q01279 mus musculu
6	3003.5	44.1	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2984	43.8	1308	1 ERB4_RAT	P62956 rattus norv
8	2717.5	39.9	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2440.5	35.8	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2369.5	34.8	1339	1 ERB3_RAT	P062799 rattus norv
11	1974	29.0	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1624	23.8	703	1 EGFR CHICK	P13387 gallus gall
16	1302	19.1	1323	1 LT23 CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculu
18	735	10.8	1363	1 ILPR BRALA	O02466 branchiosto
19	716	10.5	1382	1 INSR HUMAN	P06213 homo sapien
20	711	10.4	1383	1 INSR_RAT	P15127 rattus norv
21	710.5	10.4	1372	1 INSR MOUSE	P15208 mus musculu
22	707	10.4	1300	1 IRR MOUSE	Q9wtl4 mus musculu
23	698	10.2	1297	1 IRR_HUMAN	P14616 homo sapien
24	698	10.2	1607	1 MIPR_LYMSY	Q25410 lymnaea sta
25	693.5	10.2	1300	1 IRR_CAVPO	P14617 cavia porce
26	685	10.1	1477	1 HTK_HYDAT	Q25197 hydra atten
27	652	9.6	1367	1 IG1R HUMAN	P08069 homo sapien
28	642	9.4	1373	1 IG1R MOUSE	P060751 mus musculu
29	638.5	9.4	1370	1 IG1R_RAT	P24062 rattus norv
30	626	9.2	1390	1 INSR_ADEAE	Q93105 aedes aegyp
31	619	9.1	2146	1 INSR_DROME	P09208 drosophila
32	606	8.9	987	1 EPB4 HUMAN	P54760 homo sapien
33	591.5	8.7	984	1 EPB1_CHICK	Q07494 gallus gall

34	589.5	8.7	977	1	EPA2 MOUSE	Q03145 mus musculu
35	588	8.6	1114	1	REP2 HUMAN	P07949 homo sapien
36	584.5	8.6	976	1	EPA2 HUMAN	P29317 homo sapien
37	584	8.6	987	1	EPB4 MOUSE	P54761 mus musculu
38	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
39	579	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	FAK1 CHICK	Q00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1						
ERB2_HUMAN	ID	ERB2_HUMAN	STANDARD;	PRT;	1255	AA.
AC	P04626;					
DT	13-AUG-1987	(Rel. 05, Created)				
DT	13-AUG-1987	(Rel. 05, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)					
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell					
DE	surface receptor HER2) (MLN 19).					
GN	ERBB2 OR HER2 OR NGL OR NEU.					
OS	Homo sapiens (Human)					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86118663; PubMed=3003577;					
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,					
RA	Saito T., Toyoshima K.;					
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to					
RT	epidermal growth factor receptor.;"					
RL	Nature 319:230-234(1986).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86070181; PubMed=2999974;					
RA	Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,					
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,					
RA	Francke U., Levinson A., Ullrich A.;					
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor					
RT	shares chromosomal location with neu oncogene.;"					
RL	Science 230:1132-1139(1985).					
RN	[3]					
RP	SEQUENCE OF 737-1031 FROM N.A.					
RX	MEDLINE=86016729; PubMed=2995967;					
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;					
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the					
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a					
RT	human salivary gland adenocarcinoma.;"					
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).					
RN	[4]					
RP	VARIANTS VAL-654 AND VAL-655.					
RX	MEDLINE=93194196; PubMed=8095488;					
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;					
RT	"Characterization of a new allele of the human ERBB2 gene by allele-					
RT	specific competition hybridization.;"					
RL	Genomics 15:426-429(1993).					
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,					
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A					
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-					
CC	ALPHA AND AMPHIREGULIN.					
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein					
CC	tyrosine phosphate.					
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS					
CC	(POTENTIAL).					
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.					

QY 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKGILDEAYVMAGVGSP 780
 DB 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKGILDEAYVMAGVGSP 780
 QY 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNCWQIAKGSYLEDVR 840
 DB 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNCWQIAKGSYLEDVR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPVKWMALESILARRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLIDIDETEHADGGKVPVKWMALESILARRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMVKCWM 960
 QY 961 IDSECRPRRELVSERMRADPQRFVWIONEDLGASPLDSTFYRSLLEDMDGLDVA 1020
 DB 961 IDSECRPRRELVSERMRADPQRFVWIONEDLGASPLDSTFYRSLLEDMDGLDVA 1020
 QY 1021 EYLVFQOQGFCDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 DB 1021 EYLVFQOQGFCDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGTGVAFLTCSPOPEYV 1140
 DB 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGTGVAFLTCSPOPEYV 1140
 QY 1141 NQPDVPRPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTQP 1200
 DB 1141 NQPDVPRPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTQP 1200
 QY 1201 GGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
 DB 1201 GGAAPQHPHPAPSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
 ID_ERB2_RAT STANDARD; PRT; 1257 AA.
 AC F06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).
 DE ERBB2 OR NEU.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=86118662; PubMed=3945311;
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein.";
 RL Nature 319:226-230(1986).
 RN [2]
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=91222560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system.";
 RL Neuron 6:691-704(1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=92155181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";
 RL EMBO J. 11:43-48(1992).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS. THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 DR EMBL; X03362; CAA27059.1; ALT_INIT.
 DR PIR; A24562; TVRTNU.
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP_motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; PU; 3.
 DR SMART; SM00219; TyrKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 655 677 POTENTIAL.
 FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 159 369 CYS-RICH.
 FT DOMAIN 473 646 CYS-RICH.
 FT DOMAIN 722 989 PROTEIN KINASE.
 FT NP BIND 728 736 ATP (BY SIMILARITY).
 FT BINDING 755 755 ATP (BY SIMILARITY).
 FT ACT_SITE 847 847 BY SIMILARITY.
 FT DISULFID 196 205 BY SIMILARITY.
 FT DISULFID 200 213 BY SIMILARITY.
 FT DISULFID 221 228 BY SIMILARITY.
 FT DISULFID 225 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 312 BY SIMILARITY.
 FT DISULFID 316 332 BY SIMILARITY.
 FT DISULFID 335 339 BY SIMILARITY.
 FT DISULFID 513 522 BY SIMILARITY.
 FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.

FT DISULFID 569 586 BY SIMILARITY.
 FT DISULFID 589 598 BY SIMILARITY.
 FT DISULFID 602 625 BY SIMILARITY.
 FT DISULFID 628 636 BY SIMILARITY.
 FT DISULFID 632 644 BY SIMILARITY.
 FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

 Query Match 88.0%; Score 5994; DB 1; Length 1257;
 Best Local Similarity 87.8%; Pred. No. 2e-311;
 Matches 1104; Conservative 50; Mismatches 101; Indels 2; Gaps 2;

 QY 1 MELAAALCRWGLLALLPPGAASQVCTGTDMKRLPASPEHLDMLRHLVQGCQVQGNL 60
 DB 1 MELAAALCRWGLLALLPPGAASQVCTGTDMKRLPASPEHLDMLRHLVQGCQVQGNL 60

 QY 61 ELTYLPTNASLSFLQDIQEQGYVLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 DB 61 ELTYVPAASLSFLQDIQEQGYVLAHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNR 120

 QY 121 DPLNTPVT-GASPGRLRLQRLSLEILKGGVLTORNPOLCYQDTILWKDIFHKNQNL 179
 DB 121 DPQDNVAASPPGRTEGLRLQRLSLEILKGGVLTORNPOLCYQDMVLWKDFRKNQNL 180

 QY 180 ALTLIDNRSRACHPCSPCKGRCWSESSEDQSLTRTVACGACRCKGLPDTDCHEQ 239
 DB 181 APVDIDNRSRACHPCAPACKDNCWGESPEDQILFTICTSGCARCKRLPDTDCHEQ 240

 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMHNPEGRYTFGASCVTTC 300

 QY 300 PNYLSTDVGSCTLVCPNLNQEVTAEDGTORCKSKPCARVCYGLGMEHLREVRVTS 359
 DB 301 PNYLSTDVGSCTLVCPNNQEVTAEDGTORCKSKPCARVCYGLGMEHLRGAITS 360

 QY 360 NIOEFAGCKKIFGSLAFPLPESFDGASNTAPLOQVFELEITGLYIYSAMPDSL 419
 DB 361 NVQEFQCKKIFGSLAFPLPESFDGASNTAPLOQVFELEITGLYIYSAMPDSL 420

 QY 420 PDLVSFQNLQVIRILHNGAYSLTLQGLISWGLRLSRLRELGLSLALIHNTHLFCVHT 479
 DB 421 RDLVSFQNLRIIRGRLHNGAYSLTLQGLISWGLRLSRLRELGLSLALIHNTHLFCVHT 480

 QY 480 VPDQLFRPHQALLHTANPEDE-CVGEGLACHQICARGHCHGPGPTQCVCNCSQFLRGQ 538
 DB 481 VPDQLFRPHQALLHSGNRPEBDCVSSGLVCSNLCAGHCHGPGPTQCVCNCSHFLRGQ 540

 QY 539 ECVEECRVQLGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVA 598
 DB 541 ECVEECRVKGLPREVYSDKRLCPCHPECPQNGSVTCFGEADQCAACHYKSSSCVA 600

 QY 599 RCPGSGVKPDLSPYPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTISVAV 658
 DB 601 RCPGSGVKPDLSPYPIWKFPDEEGICQPCPINCTHSCVDLDDKCPAEQASPTFIATV 660

 QY 659 VGLLVVVLGVGGLIKRQKIRYTWRRLLQETELVEPLTPSGAMPNQAMRLKET 718
 DB 661 VGLLVVVLGVGGLIKRQKIRYTWRRLLQETELVEPLTPSGAMPNQAMRLKET 720

 QY 719 ELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAIVNAGV 778
 DB 721 ELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAIVNAGV 780

 QY 779 SPYVSRLLGLICTSTVQLTQMPYGCLLDHVRENRGLSGQDLNWCQIAKGMYSYLED 838

DB 781 SPYVSRLLGLICTSTVQLTQMPYGCLLDHVRENRGLSGQDLNWCQIAKGMYSYLED 840
 QY 839 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRR 898
 DB 841 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRR 900
 QY 899 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKC 958
 DB 901 FTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKC 960
 QY 959 WMIDSECRPRFRELSEFSEARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLV 1018
 DB 961 WMIDSECRPRFRELSEFSEARMARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDMDGLV 1020
 QY 1019 DAEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEEEAPRSLAPS 1078
 DB 1021 DAEYLVPQGGFFCPDPTEGTGTSTARRHRRSSSTRSGGDLTLGLPSEEGPRSLAPS 1080
 QY 1079 EGAGSDVFDGLGMGAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLTCSPQPE 1138
 DB 1081 EGAGSDVFDGLAMGVTKGLQSLSPHDLSPLOQYSEDPTLPLPETDGVVAPLACSPQPE 1140
 QY 1139 YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKDVFAFGGAVENPEYLT 1198
 DB 1141 YVNOSEVQPPPLTPREGPLPVVRPAGATLERPKTLPSPGNVGVKDVFAFGGAVENPEYLV 1200
 QY 1199 POGGAAPQHPPPAPSPAFDNLVYNDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 DB 1201 PREGTASPPHPSPAPSPAFDNLVYNDQNSSEQPPSPFEGTPTAENPEYLGLDVVPV 1257

 RESULT 3
 ERB2_MESAU STANDARD; PRT; 1254 AA.
 ID ERB2_MESAU Q60553;
 AC 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 GN (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; PubMed=7908275;
 RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
 RL Gene 140:251-255(1994).
 CC -! FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX;
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC -! CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -! SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -! SUBCELLULAR LOCATION: Type I membrane protein.
 CC -! PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC -! SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; D16295; BAA03801.1; -	QY	61	ELTYLPTNASLSFLQDIOEVQGYVLIHANQVQVFLQRLRIVRGTLQFEDNYVALAVLDNG	120
CC	HSSP; P11362; 1FGK.	Db	61	ELTYLPANATLSFLQDIOEVQGYVLIHANQVQVFLQRLRIVRGTLQFEDNYVALAVLDNR	120
DR	InterPro; IPR000494; EGFR_L domain.	QY	121	DPLNNTPTVYGASPGGLRELQRLSLEILKGGVLTORNPOLCYODTILWKDIFHKNQOLA	180
DR	InterPro; IPR000719; Euk_pkinase.	Db	121	DPLDNVTTATGRTPEGLRELQRLSLEILKGGVLTORNPOLCYODTILWKDIFHKNQOLA	180
DR	InterPro; IPR002174; Furin-like.	QY	181	LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSILTRTTCAGGCARCKGKPLPTDCCHEQC	240
DR	InterPro; IPR004019; YLP_motif.	Db	181	PVDIDNRSRACHPCSPMCKSGRCWGESSEDCQSILTRTTCAGGCARCKGKPLPTDCCHEQC	240
DR	Pfam; PF00069; pkinase; 1.	QY	241	AAGCTGPKHSDCLACHLHFNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTCP	300
DR	Pfam; PF00757; Furin-like; 1.	Db	241	AAGCTGPKHSDCLACHLHFNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTCP	300
DR	Pfam; PF01030; Recep_L_domain; 2.	QY	301	YNYLSTDVGSCTLVCPLNHNEVTADGTQRCCKSKPCARVCYGLGMEHLREVATSAN	360
DR	Pfam; PF02757; YLP; 2.	Db	301	YNYLSTDVGSCTLVCPLNHNEVTADGTQRCCKSKPCARVCYGLGMEHLREVATSAN	360
DR	ProDom; PD000001; Euk_pkinase; 1.	QY	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFELEETGYLYISAWPDSL	420
DR	SMART; SM00219; TyrKc; 1.	Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVFELEETGYLYISAWPDSL	420
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	QY	421	DLVSFQNLQVIRGRILHNGAYSILTLQGLISWGLRLSRLSRELGLALIHNNHLCFVHTV	480
DR	PROSITE; PS00109; PROTEIN KINASE TTR; 1.	Db	421	DLVSFQNLQVIRGRILHNGAYSILTLQGLISWGLRLSRLSRELGLALIHNNHLCFVHTV	480
DR	PROSITE; PS50011; PROTEIN KINASE DOM; 1.	QY	481	PWDQLFRLPHQALLHNTANRPEDECYVGBGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	Db	481	PWDQLFRLPHQALLHNTANRPEDECYVGBGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540
KW	Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	QY	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFCGPEADQCACAHYKDPKPCVARC	600
KW	Proto-oncogene; Disease mutation.	Db	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFCGPEADQCACAHYKDPKPCVARC	600
FT	SIGNAL 1 21 POTENTIAL	QY	601	PSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQASPLTSIVSAVVG	660
FT	CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.	Db	601	PSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQASPLTSIVSAVVG	660
FT	DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).	QY	661	ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTSGAMPNQAQMRILKETEL	720
FT	TRANSMEM 653 675 POTENTIAL.	Db	661	ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTSGAMPNQAQMRILKETEL	720
FT	DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).	QY	721	RVKVLGSGAFCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGP	780
FT	DOMAIN 158 368 CYS-RICH	Db	721	RVKVLGSGAFCTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGP	780
FT	DOMAIN 472 644 CYS-RICH.	QY	781	YVSRLLGICLTSTVOLVTQMPYGCILDHVREHRLGSGQDLNMCVQIAKMSYLEYVR	840
FT	DOMAIN 720 987 PROTEIN KINASE.	Db	781	YVSRLLGICLTSTVOLVTQMPYGCILDHVREHRLGSGQDLNMCVQIAKMSYLEYVR	840
FT	NP_BIND 726 734 ATP (BY SIMILARITY).	QY	841	LVHRLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKIMWALSILRRFT	900
FT	BP_BINDING 753 753 ATP (BY SIMILARITY).	Db	841	LVHRLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVPKIMWALSILRRFT	900
FT	ACT_SITE 845 845 BY SIMILARITY.	QY	901	HOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKVM	960
FT	DISULFID 195 204 BY SIMILARITY.	Db	901	HOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWKVM	960
FT	DISULFID 199 212 BY SIMILARITY.	QY	961	IDSECRPRELVSFSSRMARDPQRVFIQNEBGLDGPASPLDSTFYRSLLDDMDGLVDA	1020
FT	DISULFID 236 244 BY SIMILARITY.	Db	961	IDSECRPRELVSFSSRMARDPQRVFIQNEBGLDGPASPLDSTFYRSLLDDMDGLVDA	1020
FT	DISULFID 240 252 BY SIMILARITY.	QY	1021	ESYLVQQGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPASPSAPSG	1080
FT	DISULFID 255 264 BY SIMILARITY.	Db	1021	ESYLVQQGFFCPDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEPASPSAPSG	1080
FT	DISULFID 268 295 BY SIMILARITY.	QY	1081	AGSDVFDGDLGMAAKQLSLPHTDPSPLORYSEDTVPLPSETGYVAPLTCSPQEVV	1140
FT	DISULFID 299 311 BY SIMILARITY.	Db	1081	AGSDVFDGDLGMAAKQLSLPHTDPSPLORYSEDTVPLPSETGYVAPLTCSPQEVV	1140
FT	DISULFID 315 331 BY SIMILARITY.				
FT	DISULFID 334 338 BY SIMILARITY.				
FT	DISULFID 511 520 BY SIMILARITY.				
FT	DISULFID 515 528 BY SIMILARITY.				
FT	DISULFID 531 540 BY SIMILARITY.				
FT	DISULFID 544 560 BY SIMILARITY.				
FT	DISULFID 563 576 BY SIMILARITY.				
FT	DISULFID 567 584 BY SIMILARITY.				
FT	DISULFID 587 596 BY SIMILARITY.				
FT	DISULFID 600 623 BY SIMILARITY.				
FT	DISULFID 626 634 BY SIMILARITY.				
FT	DISULFID 630 642 BY SIMILARITY.				
FT	MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).				
FT	MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).				
FT	CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).				
FT	VARIANT 658 658 V -> E (IN ONCOGENIC NEU).				
FT	VARIANT 659 659 V -> E (IN ONCOGENIC NEU).				
SQ	SEQUENCE 1254 AA; 138252 MW; 974C3791C21P2BEI CRC64;				

Query Match 87.9%; Score 5984.5; DB 1; Length 1254;
 Best Local Similarity 87.6%; Pred. No. 6.3e-311;
 Matches 1099; Conservative 58; Mismatches 97; Indels 1; Gaps 1;
 1 MELAAACRWGLLALLPPGNASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVQVQGNL 60
 1 MELAAACRWGLLALLSPGASGTQVCTGTDMLRLPASPTHLDIVRHLYQGCVQVQGNL 60

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts."
 RL Growth Factors 13:121-132 (1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT An-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor."
 RL J. Biochem. 127:65-72 (2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Oda M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor."
 RL J. Biol. Chem. 273:11150-11157 (1998).
 RN [19]
 RP REVIEW
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens."
 RL Annu. Rev. Biochem. 56:881-914 (1987).
 CC -I- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -I- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -I- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -I- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 Query Match 46.5%; Score 3166; DB 1; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 4.3e-161;
 Matches 630; Conservative 177; Mismatches 352; Indels 106; Gaps 21;
 QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPEHLDMLRHLYOGCQVQGNLELTLPFN 69
 DB 14 LLAALCPASRALEBKVKCQTSNKLTLQGLTFEDHFLSLQRMFNNECVLGNLEITYVRN 73

QY 69 ASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLTKIQEVAGYVLIALTNTVERIPLENLQIRGNMYVENSALAVLSNYD----- 136
 QY 129 VTGASPGGLREQLRLSLTILKGGVLIQRNPOLCYODTILWKDIFKHNQLALTLDITNR 188
 DB 127 ---ANKTGLKELPMRNLQELHGAVRFSNNPALCNVESIQWRDIVSDSFLSNMDFQNH 183
 QY 189 SRACHPCSPCKSGRCWGSSESDCQSLTRTVACAGCA-RCKGGLPTDCHEQCAAGCTGP 247
 DB 184 LSCQCKDPCSPNGSCWGAEECNCKLTKIIQAQCSGRCRKGKSPSCDCHNCAAGCTGP 243
 QY 248 KHSDDLACLHFNHSGICELHCPALVYNTTDTFESMPNPEGRYTFGASCVTACPNYLSLD 307
 DB 244 RESDCLVCRKFRDEATCKDTPPLMLNPTTYQMDVNPGEKYSFGATCKVKCPRNVTVD 303
 QY 308 VGSCTLVCPLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLREVRVTSANIQFAGC 367
 DB 304 HGSVRAACGADSYEM-EEDGVKCKCKGCGRKVCNGIGIGEFKDSLSINATNIKFKNC 362
 QY 368 KKFISLAFLPESFUDGPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLPLDLVFQON 427
 DB 363 TSSISGDLHLPLVAFRGDSFTHTPPDQBLDILKTVKEITGFLLOAWFENRTDLHAFEN 422
 QY 428 LQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGSGLALIHNNHLCFVHTVPMQDLFR 487
 DB 423 LEIRGRTKQHGQFSLAVVSLNITSLGLSLKLEISDGDVVISGNKNLCYANTINWKKLFG 482
 QY 488 NPHQALLHTANPDECEGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQECVSCRVL 547
 DB 483 TSGQTKIISNRENSCKATGQVCHALCSPEGCGPEPRDCVSCRNVSGRECVDCNLL 542
 QY 548 QGLPREYNARHCLCPHCPCQNGSVTCFGEADOCVACAHKDPFCVARGCPSGVKPD 607
 DB 543 EGEPRFVENSICIQHPCLQAMNITCTGRGPDNCIOCAHYIDGPHCVKTCPCAGVMGE 602
 QY 608 LSYMPIWKFPEBEGACQPCFINCTHSCVDLDKGCFAEQRASPLTSIVSAVVG---ILLV 664
 DB 603 NNTL-VKMYADAGHVCHLCHPNCTYCTGPGLEGCTNGPKIP--SIATGMVGAALLLV 659
 QY 665 VLVGVVGLIIRROOKIRKYMRRLLOETELVEPLTPSGAMPNQAOIRLKEETURKVK 784
 DB 660 VALGIG---LFWRRRRIHVRKTLRRLQERLEVEPLTPSGEAPNQALLRLKETEFKKIK 716
 QY 725 VLGSAGFTVYKGIWTPDCENVKIPIVAIKVLRNTSPKANKELDEAYVMAGVGSVYVR 784
 DB 717 VLGSAGFTVYKGLWIPGEKVKIPIVAIKELREATSPKANKELDEAYVMASVDNPHVCR 776
 QY 785 LLGICLTSTVQLVTQIMPYGCLLDHVRENRGRGLSGQDLNLCWQIAKGMYSYLEDVRLVHR 844
 DB 777 LLGICLTSTVQLITQLMPFGCLLDVREHKDNIGSQYLLNWCQIAKGMNYSLEDRLVHR 836
 QY 845 DLAAARNLVKSNHVKITDFGLARLLDDDETHADGGKVPKWMMALESILRRRFTHQSD 904
 DB 837 DLAAARNLVKTPQHKITDFGLAKLLGAEKEHYAEGGKVPKWMMALESILHRIYTHQSD 896
 QY 905 VMSYGVYVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYMWKCMWIDSE 964
 DB 897 VMSYGVYVWELMTFGSKPYDGIPIASEISILEKGERLPQPICTIDVYMWKCMWIDAD 956
 QY 965 CRPRFRELVSFPRMARDPQRFWVIQ-NEDLGPASPLDSTFYRSLLEDMDGDLVDAEY 1023
 DB 957 SRPKFRELIIEFKWARDPQRLVLVQGDERMHLPSFTDSNFYALMDEEDMDVDVADY 1016
 QY 1024 LVPQCGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSGAGS 1083
 DB 1017 LIPQCGFF-----SSPSTSRTPLLSSLSATS 1042
 QY 1084 DYFDGDLGMAAKGLQSLPHTDPSPLOYSEDPVTPLPSET--DGYVAPLTCSPQEPYVN 1141
 DB 1043 N--NSTVACIDRNLGQSCPIKEDSFLQSYSSDPTGALTEDSIDDTFL-----PVPEYN 1094

```
QY 1142 QDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKVDFAFGAVENPEYL-TPQ 1200
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1095 Q-SVPRKPAQSVQNVYHNOPLNP-----APSRDPHYQD--PHSTAVCNPEYLNWQ 1143
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1201 GGAAPQHPHPPAFSPADNLYYWDQ-----DP-----PERGAPSTFKGPTTAE 1244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1144 -----PTCVNSTFDSPAHWAQKSHQISLDNPDYQDQDFPKKAPNGIFKGS-TAE 1193
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1245 NPEYL 1249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
(bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetkeke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Bisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
(PHOSPHORYLATION (BY PKC) (BY SIMILARITY)).
PHOSPHORYLATION (AUTO-) (BY SIMILARITY)).
```

```
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA5587.1; -
CC EMBL; U03425; AAA1789.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSSP; P11362; IFGK.
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
CC DOMAIN 25 647 POTENTIAL.
CC TRANSMEM 648 670 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
CC REPEAT 75 300 APPROXIMATE.
CC REPEAT 390 600 APPROXIMATE.
CC DOMAIN 1028 1071 SER-RICH.
CC NP_BIND 720 728 ATP (BY SIMILARITY).
CC BINDING 747 747 ATP (BY SIMILARITY).
CC ACT_SITE 839 839 BY SIMILARITY.
CC DISULFID 190 199 BY SIMILARITY.
CC DISULFID 194 207 BY SIMILARITY.
CC DISULFID 219 223 BY SIMILARITY.
CC DISULFID 232 240 BY SIMILARITY.
CC DISULFID 236 248 BY SIMILARITY.
CC DISULFID 251 260 BY SIMILARITY.
CC DISULFID 264 291 BY SIMILARITY.
CC DISULFID 295 307 BY SIMILARITY.
CC DISULFID 311 326 BY SIMILARITY.
CC DISULFID 329 333 BY SIMILARITY.
CC DISULFID 506 515 BY SIMILARITY.
CC DISULFID 510 523 BY SIMILARITY.
CC DISULFID 526 535 BY SIMILARITY.
CC DISULFID 539 555 BY SIMILARITY.
CC DISULFID 558 571 BY SIMILARITY.
CC DISULFID 562 579 BY SIMILARITY.
CC DISULFID 582 591 BY SIMILARITY.
CC DISULFID 595 617 BY SIMILARITY.
CC DISULFID 620 628 BY SIMILARITY.
CC DISULFID 624 636 BY SIMILARITY.
CC MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
 (BY SIMILARITY).
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 19 19 C -> S (IN REF. 2).
 FT CONFLICT 539 539 C -> W (IN REF. 5).
 FT CONFLICT 991 991 L -> F (IN REF. 4).
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46D2D2F5 CRC64;
 Query Match 46.2%; Score 3145; DB 1; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 5.7e-160;
 Matches 633; Conservative 170; Mismatches 359; Indels 110; Gaps 23;
 QY 11 LLLALLPPGAA--STQVCTGTDKMLRLPASPEHLDMLRLHYOGCQVQGNLELTYPTN 68
 DB 14 LLLTALCAAGALBEKKVCOGTNRLLTQGTGFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73
 QY 69 ASLSFLQDIOEVQGYLIAHNOVQVPLQRLIRVRGTQOLFEDNVALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLQIRGNALYENYVALAILSN----- 124
 QY 129 VTGASPGGLRELQRLSITEILKGGVLIQRPOLCYQDTILWKDI-----FKNNQLALTLI 184
 DB 125 -YGNRTGLRELPMNLQELIAGVRSNNPILCNMDTIQWRDVIQNVFNSMSDL--- 180
 QY 185 DTRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVACGGCA-RCKGPLPTDCHEQCAAG 243
 DB 181 -QSHSPCKPCDFPCNCGWCGGEENCCQLTKIICAAQCQSHRCGRSPSDCCNCAAG 239
 QY 244 CTGPKHSDDLACIENHSGICELHCPALVTYNTDTESMPNPEGRVTFEGASCVTAPYNY 303
 DB 240 CTGPRSDCLVCKQFQDEATKDTCTPLMLYNTTYQMDVNPBGKVSFGATVKKCPNY 299
 QY 304 LSTDVGSCTLVCLHNOQVTAEDGTORCEKSCPCARVCYGLGHEHLREVRVAVTSANIÖE 363
 DB 300 VVTDHGSVCRAAGPDYEV-EEDGIRKCKDCGCRKVCNGIGIGEFKDTLSINATIKH 358
 QY 364 FAGCKIFGSLAFPLPSFDGDPASNTAPLOEQVFPETLEETGYLYISAMPDLSPLDLS 423
 DB 359 FKCYTAISGDLHLPVAFKGDSTRTPLDPRELEILKTVKEITGFLIIQAWPDNWTDLH 418
 QY 424 VFONLQVIRGRIHLNGAYSILTLOGLSISGLSLRELGLSLALIHNTLHLCFHVTPWD 493
 DB 419 AFENLEIRTKQHGQFSLAVVGLNLTSLGLSLKEISDGVIIISGNRLCYANTINWK 478
 QY 484 QLFERNPHQALLHTANRPEDECVGEGLACHCARGHCWGPQTCVNCISOFLRGQECVEE 543
 DB 479 KLFGTNPQTKIWNNAEAKDCKAVHVCNPLCSGECGCEPDRDCVQNVSGRECEVK 538
 QY 544 CRVLOGLPRYVNAHCLPCHPCQPNQSVTCFGEADQCVACAHYKDPFPCVACPSG 603
 DB 539 CNLIEGEPREFVENSEICIQHCEPLQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCAPG 598
 QY 604 VKPDLSPMTWPKFDEGACQPCPINTHSCVDLDDKGCFAEQRASPLTISVSAVVGILL 663
 DB 599 IMGENNTL-VWKYADANNVCHLCHANCTYACAGLQGCVEWSPGPKIPSIATGIVGGL 657
 QY 664 VVVLGVVFGI-LIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAOMRIKTELK 722
 DB 658 FIVV-VALGIGLFWRRHRIKVTLLRLLQRELVEPLTPSGAPNQAHLRIKTEFKK 716
 QY 723 KVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPTKANKEILDEAYVMAGVSPY 782

Db 717 IKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPTKANKEILDEAYVMASVDNPHV 776
 QY 783 SRLIGCLTSTVOLVTQMLPYPGCLLDHVRNRRGLSQDILLNMCWQIAKMSYLEVDRLV 842
 Db 777 CRLLGCLTSTVOLITQMLPYPGCLLDYVREKDNIGSQVLLNWCWQIAKGMVLEDRLV 836
 QY 843 HRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWMALLESILRRRPTHQ 902
 Db 837 HRDLAARNVLKTPQHKITDFGLAKLLGAEKEEYHAEKGVPIKWMALLESILHRIYTHQ 896
 QY 903 SDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPORPPICTIDVYIMVCKMID 962
 Db 897 SDVMSYGVTVWELMTGSKPYDGPASDISILEKGERLPORPPICTIDVYIMVCKMID 956
 QY 963 SECPRFRRLVSEFSRMARDPQRFVVIQ-NEDLGPASPLDSTFYRSILLEDDDDGLVDAAE 1021
 Db 957 ADSRPFERLILFESRMARDPQRYLVIQDERMHLSPDTSNFYRALMBEEDMEDVVDAD 1016
 QY 1022 EYLVPQOGFFCPDPAPGAGMVHRRSSSTRSGGGDLTLGLEPSEEAAPRSPAPSEGA 1081
 Db 1017 EYLVPQOGFF-----NSPST-----SRTELLSSLSA 1042
 QY 1082 GSDVFDGDLGMAAKGLOSLPHTDPSLPQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
 Db 1043 TSN-----NSTVACINRNGSCRKEDAFQRYSDPTGAVTEONIDDAFL-----VPPEY 1092
 QY 1140 VNOPDVRPOPSPREGPLPAAPAGATLERAKTTLSPKNGVVKDVAFGAGAVENPEYL-T 1198
 Db 1093 VNO-Q-SVPKPPASQVNPVHNQPLHP-----AFGRDLHYQN--PHSNVAGNPEYLVNT 1141
 QY 1199 PQGGAAPPHPPAFSPAFDNLYYWDQ-----DP-----PERGAPSPFTKGTPT 1242
 Db 1142 AQ-----PTCLSSGGENSPALWIKQSHQWMLNDPNYQDDFFPKETKPNKIFKG-PT 1191
 QY 1243 AENPEYLGIDVP 1254
 Db 1192 AENAEYLRVAPP 1203
 RESULT 6
 ERB4_HUMAN STANDARD; PRT; 1308 AA.
 ID ERB4_HUMAN AC Q15303;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
 DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
 GN ERB4 OR HER4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).
 RC TISSUE=Breast carcinoma;
 RX MEDLINE=93189574; PubMed=8383326;
 RA Plowman G.D., Culouscou J.-N., Whitney G.S., Green J.M., Carlton G.W.,
 RA Foy L., Neubauer M.G., Shoyab M.;
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
 RT epidermal growth factor receptor family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
 RC TISSUE=Fetal brain;
 RX MEDLINE=97476287; PubMed=9334263;
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
 RA Klagsbrun M.;
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
 RT tissue distribution and differential processing in response to
 RT phorbol ester.";
 RL J. Biol. Chem. 272:26761-26768(1997).


```
Db      536  CNLYDGEFFREFNGSICVECDPOCEKXMEDGLTCHGPGDNCTKSHFKDGNPCVCKPD 595
Qy      603  GVKPDLISYMPIWPKFPEEGACQCPINCTHSCVDLDDKGC-----PABQASPL 651
Db      596  GLQCANSF--IFKYADPDRECHPCPNCTQCGNPTSHDCIYYPWTGSHSTLPQHAR-TPL 652
Qy      652  TSTVSAVV-CILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQA 710
Db      653  --TAAGVIGGLFILLVIGLTFVYVRKRSIK-KRALLRRL-ETELVEPLTPSGTAPNQA 708
Qy      711  QMRILKETELRKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770
Db      709  QLRILKETELRKVKVLSGAGFYVYKGIWIPDGETVKIPVAIKILNETTGPKANVFEMDE 768
Qy      771  AYVAGVGSYVSRLLGICLTSTVLQVLTQMPYGCCLLDHVRNRLGSGDOLLNWCQVIA 830
Db      769  ALIMASMDHPLVRLLLGCVLSPTIQLVTLMPHGLLEYVHEHKDNIGSQLLLNWCQVIA 828
Qy      831  KGMSTLEDVLRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKQWMA 890
Db      829  KGMSTLEERLVRHDLAARNVLKSPNHVKITDFGLARLLGDEKEYNADGGKMPKQWMA 888
Qy      891  LESILARRFTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTID 950
Db      889  LECIHVRKFTHQSDVMSYGVYVWELMTFGKPYDGIPTREIPDLLEKGERLPQPPCTID 948
Qy      951  VYIMVKCMWIDSECRPRFRELVSFSRMDARDQRFVVIQNEP-LGFPASPLDSTFYRSL 1009
Db      949  VYIMVKCMWIDASRPKPKELAAEFSRMDARDQRFVVIQNEPDRMKLPSPNDSKFFQNLL 1008
Qy      1010  EDDMGDLVDAEYLPVQGFPCDPAPGAGMVHHRSSSTRSGGDLTLGLEPSEEE 1069
Db      1009  DEEDMGDLVDAEYLPV-QAFNIPP-----IYTSRARIDSRS-----EIGHSPPPAY 1056
Qy      1070  APS-----PLAP--SEGAGSDVFDGLGMAAGLQ 1100
Db      1057  TPNMGQFVYRDCGFAAEQGVYVPYAPTSTIEAPVAQATAEIFDDSCNCTLKRKPA 1116
Qy      1101  LPTHPSPLQRYSEDTVPPLPS-----ETDGYAPLTCSPQPEYVNPQDVRPQPPSPR 1153
Db      1117  PHVQEDSSTQRYSDPTVPAPERSPRGELDEEGVMTMRDKPKQVYLNVE----- 1167
Qy      1154  EGPLPAPRPAATLERAKTLPSCNGVGVKDVFAFGGAVENPEYLTPOGGAAPQHPPPA- 1212
Db      1168  ENPFVSR-----KNGDLQ-----ALDNPYHNASNG-----PPKAE 1199
Qy      1213  -----FSPAFONLYYWDQDPPPERGA--PPSTF 1237
Db      1200  DEYVNEPLYNTFANTLGKAEYKNNILSMPEKAKKAFDNPDYWNHSLPRSTLQHPDYL 1259
Qy      1238  KGTF-----AENPEYL 1249
Db      1260  QEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID      ERB4_RAT  STANDARD; PRT; 1308 AA.
AC      O62956; Q922N7;
DT      15-DEC-1998 (Rel. 37, Created)
DT      15-JUN-2002 (Rel. 41, Last sequence update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN      ERB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Heart;
RX      MEDLINE=98221155; PubMed=9553078;
RA      Zhao Y.-Y., Sawyer J.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes.";
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      [2]
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=91222560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system.";
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration.";
RL      J. Neurosci. 17:1642-1659(1997).
CC      -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AF041838; AAD08899.1; -.
CC      EMBL; U52531; AAC53051.1; -.
CC      HSSP; P11362; 1FGK.
CC      InterPro; IPR000494; EGFR_L domain.
CC      InterPro; IPR000719; Euk_pkinase.
CC      InterPro; IPR002174; Furin-like.
CC      InterPro; IPR001245; Tyr_pkinase.
CC      InterPro; IPR004019; YLP motif.
CC      Pfam; PF00757; Furin-like; 1.
CC      Pfam; PF00069; pkinase; 1.
CC      Pfam; PF01030; Recep_L domain; 2.
CC      Pfam; PF02757; YLP_2.
CC      PRINTS; PR00109; TYRKINASE.
CC      ProDom; PD000001; Euk_pkinase; 1.
CC      SMART; SM00261; FU; 4.
CC      SMART; SM00219; TyrKc; 1.
CC      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE; PS00111; PROTEIN_KINASE_TYR; 1.
CC      PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
CC      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC      Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
CC      SIGNAL 1 25 POTENTIAL.
CC      CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
CC      DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
CC      TRANSMEM 652 675 POTENTIAL.
CC      DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```

FT	DOMAIN	186	334	CYS-RICH.	QY	357	TSANIOEAGCKKIFGSLAFPESEFDDPSASNTAPLQPEQLQVFELEBITGYLISAWP	416
FT	DOMAIN	496	633	CYS-RICH.	Db	349	SSNDKFNCTKINGNLIFLVTHGIDPYNDAIDPEKLNVRVTRITGLNTQWP	408
FT	NP BIND	724	732	ATP (BY SIMILARITY).	QY	417	DSLPLDSVFNQVIRGRILHNGAYSLTLQGLISWGLRSLRSLGSLALIHNNHLCF	476
FT	BINDING	751	751	ATP (BY SIMILARITY).	Db	409	PNMTDFSVSNLVTIGGRVLSGLSLILKQOQITSLQFQSLKEISAGNIYITDNLNLCY	468
FT	ACT SITE	843	843	BY SIMILARITY.	QY	477	VHTVPWDLFRPHQALHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCSQFLR	536
FT	DISULFID	189	197	BY SIMILARITY.	Db	469	YHTINWTTLPSTVNQIRIVDRNRAENCTAEGVMVNCNLDGCGPGDPDCLSCRRFSR	528
FT	DISULFID	213	221	BY SIMILARITY.	QY	537	GOECVEECRVLOGLPREYVNAHCLPCHPECOF-QNGSVTCFPEADQCACAHYKDPFF	595
FT	DISULFID	217	229	BY SIMILARITY.	Db	529	GKICIESCNLYDGEFEPENGSIQVECDQCEKMDGLTCHGPGDNCCKSHFKDGN	588
FT	DISULFID	230	238	BY SIMILARITY.	QY	596	CVARCPGKVPDLVSYMPIWKFPDEBAGCQPCINCTHSCVDLDDKGC-----PA	644
FT	DISULFID	234	246	BY SIMILARITY.	Db	589	CVEKCPDLQGANSF--IFKYADQDRECHPCPNCTQGCNGPTSHDCIYYPWTGHTLPQ	646
FT	DISULFID	249	258	BY SIMILARITY.	QY	645	EQASPLTISVSAV--GILLVVVLGVVGLIKRQOKIRKYTMRELLLOETELVEPLTPS	703
FT	DISULFID	262	289	BY SIMILARITY.	Db	647	HAR--TFL--IAAGVIGGLFILVIMALTFAVYVRRSIK-KKRLRRFL--ETELVEPLTPS	701
FT	DISULFID	293	304	BY SIMILARITY.	QY	704	GAMPNQAQMRILKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVRENTSPKA	763
FT	DISULFID	308	323	BY SIMILARITY.	Db	702	GTAPNQAQRLIKETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVRENTSPKA	761
FT	DISULFID	326	330	BY SIMILARITY.	QY	764	NKEILDEAYMAGVSGPYVSRLLGICLTSTVOLVQLMPYGCILLDHVRENRLGSDQLL	823
FT	DISULFID	503	512	BY SIMILARITY.	Db	762	NVEFMDALIMASVDHPLVRLVGLVCLSPITQVLTQMPHGCLLEYVHEHKONIGSOLL	821
FT	DISULFID	507	520	BY SIMILARITY.	QY	824	NMCQIAKMSVLEVDVLRDLAARNVLKSPNHVKITDFGLARLLDIDEYEHADGGK	893
FT	DISULFID	523	532	BY SIMILARITY.	Db	822	NMCVQIAKMGMYLEERLVRDLAARNVLKSPNHVKITDFGLARLLDIDEYEHADGGK	891
FT	DISULFID	536	552	BY SIMILARITY.	QY	884	VPIKMALESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ	943
FT	DISULFID	555	569	BY SIMILARITY.	Db	882	MPKMALESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ	941
FT	DISULFID	559	577	BY SIMILARITY.	QY	944	PITCTIDVIMVVKMIDSECRPRFRELVSFBSRMARDPQRFVIVQNEED-LGPASPLDS	1002
FT	DISULFID	580	589	BY SIMILARITY.	Db	942	PITCTIDVIMVVKMIDSECRPRFRELVSFBSRMARDPQRFVIVQNEED-LGPASPLDS	1001
FT	DISULFID	593	614	BY SIMILARITY.	QY	1003	TFVRSILLEDDMGDLVDAEYLVPQGGFCPPD-----APGA	1039
FT	DISULFID	617	625	BY SIMILARITY.	Db	1002	KFFQNLLEDEEDLMDMAEYLVP-QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTPMS	1060
FT	DISULFID	621	633	BY SIMILARITY.	QY	1040	GMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEAGSDVDFGDLGWAAGLQ	1099
FT	DISULFID	621	633	BY SIMILARITY.	Db	1061	GSQFVYDGGFATQGG--MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNGTLRKPV	1115
FT	DISULFID	621	633	BY SIMILARITY.	QY	1100	SLPHTDPSLPQYSEDPVTPLPS-----ETDGVYVAPLTCSPQPEYVQNPVVRPQPPSP	1152
FT	DISULFID	621	633	BY SIMILARITY.	Db	1116	VPHVQEDSSTQYSDADPTVFAPERNPRABLDSEGYMTMHDKPKQBYLNPVE-----	1167
FT	DISULFID	621	633	BY SIMILARITY.	QY	1153	REGPLPAARPACATLERAKTLSPGKGVVVKVFAFGAVENPEYLTPOGGAAPQHPHPPA	1212
FT	DISULFID	621	633	BY SIMILARITY.	Db	1168	-ENPFVSR--KNGDLQ-----ALDNPYHSASSG-----PPKA	1198
FT	DISULFID	621	633	BY SIMILARITY.	QY	1213	-----FSPAFDNLNYYDQDPPERGA--PPST	1236
FT	DISULFID	621	633	BY SIMILARITY.	Db	1199	EDEYVNEPLYNLFTNALGNAEYMKNSLLSVEKAKAFDNDPYNHSLPPRSTLQHPDY	1258
FT	DISULFID	621	633	BY SIMILARITY.	QY	1237	FKGTPT-----AENPEYL	1249
FT	DISULFID	621	633	BY SIMILARITY.	Db	1259	LOEYSTKYFYKQNGRIRPIVAENPEYL	1285

RESULT 8

XMRK_XIPMA

ID XMRK_XIPMA

AC P13388;

PRT; 1167 AA.

Query Match 43.88; Score 2984; DB 1; Length 1308;

Best Local Similarity 45.48; Pred. No. 2.2e-151; Matches 612; Conservative 191; Mismatches 384; Indels 160; Gaps 28;

QY	1	MELA-ALCRWGLLL--ALLPFGAASCTQCTGTMDKRLRLPASPTHLDMLRHLRYGQGVVQ	57
Db	1	MKLATGLVWVGLLVAARTVQPSASQSVCASTENKLSLSLDLEQQVYALRYKYECEVM	60
QY	58	GNLELTVLPNASLSFLDIOEVQGVYVLIHANNQVQVLPQRLRIVRGTLQFEDNYALAVL	117
Db	61	GNLEITSIEHNRDLSPFLRSIREVTGYVVALNQFVLPLENRAIRIGTKLYEDRYALAI	120
QY	118	DNGDPLNNTTPTVTCASPGGLRELQLRSLTEILKGGVLIQNPOLCYQDITLWKDIHKN	177
Db	121	LNTRKQGNF-----GLQELGKLNLTILNGVYVQDQKFLCYADTIHWQDIVRNPW	171
QY	178	QLALTILIDTRSRACHPCSPMKGRCWGSSESDCSLTTRVCAGGC-ARCKGPLPTDCC	236
Db	172	PSNMTLVSTIGSGGRCRCHSKCTG-RCWGPTEHNCQTILFTVCAEQDGRCYGYVSDCC	230
QY	237	HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV	296
Db	231	HRECAGSGCKPDTDCFACMNFNDSGACVTCQFQTFVYNPTTTFQLEHNFNAKYTYGAFV	290
QY	297	TACPNYLSLTVGSCSLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRV	356
Db	291	KKCPHNFV-VDSSECRACPSKMEV--EENGIRKCKPCTDIDCPKACDGTGTSLSMAQTV	348

DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
 RA Telling A., Robertson S.M., Scharf M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421 (1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharf M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X16891; CAA34770.2; -.
 DR PIR; S06142; S06142.
 DR HSP; P11362; IFGK.
 DR InterPro; IPR000494; EGF_R_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TyKc; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT POTENTIAL.
 FT DOMAIN 666 1167
 FT CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 710 977
 FT ATP (BY SIMILARITY).
 FT BINDING 716 724
 FT ACT_SITE 743 743
 FT ACT_SITE 835 835
 FT BY SIMILARITY.
 FT DISULFID 195 204
 FT BY SIMILARITY.
 FT DISULFID 199 212
 FT BY SIMILARITY.
 FT DISULFID 220 228
 FT BY SIMILARITY.
 FT DISULFID 224 236

FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 311 BY SIMILARITY.
 FT DISULFID 315 330 BY SIMILARITY.
 FT DISULFID 333 337 BY SIMILARITY.
 FT DISULFID 504 513 BY SIMILARITY.
 FT DISULFID 508 521 BY SIMILARITY.
 FT DISULFID 524 533 BY SIMILARITY.
 FT DISULFID 537 553 BY SIMILARITY.
 FT DISULFID 556 569 BY SIMILARITY.
 FT DISULFID 560 577 BY SIMILARITY.
 FT DISULFID 593 615 BY SIMILARITY.
 FT DISULFID 618 626 BY SIMILARITY.
 FT DISULFID 622 634 BY SIMILARITY.
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 201 201 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 365 365 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 398 398 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 417 417 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 621 621 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;
 Query Match 39.9%; Score 2717.5; DB 1; Length 1167;
 Best Local Similarity 45.5%; Pred. No. 2.9e-137;
 Matches 577; Conservative 163; Mismatches 390; Indels 137; Gaps 27;
 QY 4 AALCGMLLLALLPGGAAT---QVCTGDMKRLPASPETHDLMLRHLQCGQVVQGN 59
 DB 8 ALLQ--LLVLISRCSTDPDRKVCQTSNQMT---LDNHYLKKMKWYSCNVLEN 62
 QY 60 LELYPTNASLFLQDIOEVQGYVLIANHVRQVPLQRLIRIVRGQTQLFEDNYVALVDN 119
 DB 63 LEITYTQENQDLFLQSIQEVGQVLIAMNEVSTIPLVNLRLIRGQNLVEGFTLVMSN 122
 QY 120 GPLNNTTPTVTGASPGRLRLQLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNQL 179
 DB 123 YQK-NESSP--DVYQVGLQQLSLNTEILSGGVKVSHPNPLCNVETINWMDIVDKTSNP 179
 QY 180 ALTIDITNRSRACHPCSPMKSGSRCSGSESDQSLTRTVCAAGC--ARCKGPLPTDCCH 238
 DB 180 TYNLIPHAPEQCOKCDHCVCVNSCWAPGPGHCQKFKLLCAEQCNRCRGPRPICCNE 239
 QY 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYVNTDTFESMPNPEGRYTFGASCVTA 298
 DB 240 HCAGGCTGPRATDCLACRDFNDGTCCKTCTPPPKIYDIVSHQVVDNPNIKYTFGAACVKE 299
 QY 299 CPYNYLSTDVGSCTLYVCPHNOEVTAEQTCRCKSKPCARVCYVCLGMEHLREVAVTS 358
 DB 300 CPSNYVYTE--GACVRSKCSAGMLEVD--ENGRSKCPDGVCPKVCDDGIGLSLNTIAVNS 357
 QY 359 ANIQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFTLEIEITGLYVISAPDS 418
 DB 358 TWIRSESNCTKINGDIIILNRNSFEGDPHYKIGTMDPEHLWNLTTVKIEITGLYVIMWPEEN 417
 QY 419 LPDLSVFQNLQVIRGRIHLNGAYS--LTLOGLGSIWGLSLRSLRELGSGLALIHHTLCFV 477
 DB 418 MTSLSVFQNLRIIRGRTTTSRGFSFVVQVVRHLQWLGLSLRSLKEVSAGNVILKNTLQRYA 477
 QY 478 HTVPMDQLFRNPHQALLHTANRPEBCEVGEGLACHOLCARGHCWGPGPTQCVNCSQFLRG 537
 DB 478 NTINWRRLFRSEDSQIEYDART-----ENQTCNNECEDGCGWGPPTMCVSLHVRG 530
 QY 538 QECVEECRVLOGLPREYVNNARHCLPCHPRCPQNGSVTCFGEADOCVACAHYKDPFCV 597
 DB 531 GRCAVSCNLTQGEPREAQVDGRVCVQCHQECQLVQTSLTCTYGPGPANCSKSAHFQDGPQCI 590
 QY 598 ARCPGKVPDLSPYMPIWKFPDEBEGACQPCPINCTHSCVDLDDKGCQPAEQASPLTIVSA 657

```
Db 591 PRCPHGLGDDTL-INKYADKMGQCPCHQNCQTQCGSPGLSCRGD-IVSHSLAVGL 648
Qy 658 VVGILLVVLGVFGIILKRRQKIRKYVRRLLQETELVEPLTPSGAMNQAMRLIKE 717
Db 649 VSGLLITVALLVLLVLLRRRIK-RKRTIRCLQEKELVEPLTPSGOAFNQLRLIKE 707
Qy 718 TELRKVVLGSGAGFTYKGIWIPDGENVKIPVAKVLRNTPKANKETLDEAYVMAGV 777
Db 708 TEFKORVLGSGAGFTYKGLWPDGENIRIPVAKVLRNTPKANKETLDEAYVMAGV 767
Qy 778 GSPVSVRLGICLTSTVOLVQLMPYGCCLLDHVRNRRGLSQDLNWCNQIAKMSYLE 837
Db 768 DHPHVCRLGICLTSAVOLVQLMPYGCCLLDYVRQHQERCGQWLLNWCQIAKMSYLE 827
Qy 838 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRR 897
Db 828 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGKGVPIKMALESILQW 887
Qy 898 RFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDYVMIMVK 957
Db 888 TYHQSDVWSYGVTVWELMTFGSKPYDGIPIAKEIASVLENGERLPQPPICITIEVYMIILK 947
Qy 958 CWMIDSECRPRRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDODMGDL 1017
Db 948 CWMIDPSRRPRRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDODMGDL 1002
Qy 1018 VDAEYVLVPOQGFPCPDPAAGGWHHRSSSTRSGGDLTLGLEPSEEAAPRPLAP 1077
Db 1003 VDAEYVLVYKRI-----NRQGS-----BFCIP 1025
Qy 1078 SEGAGSDVFDGLGMAAGKGLQLPHTDPSPLORYSEDPTV-PLPSTDCGVAPLTCSPQ 1136
Db 1026 PTH-----PVRENSITURNISDPTQNALEKDLQGH----- 1056
Qy 1137 PEYVNQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGVVKD 1183
Db 1057 -EYVNQPGSTSSRLSDIYNPNVEDLTDGNGVPSLSQEAETNFSREYILNTQNSL--- 1112
Qy 1184 VFAFGGAVENPEYLTPOGGAAPQHPAPAFDNLYYWDQDPPERGAPPSTFKGTPTA 1243
Db 1113 PLVSSGSMDDPDY---QAG-----YQAAP-----LPQTGALTNGMFLPAA 1150
Qy 1244 ENPEYLG 1250
Db 1151 ENLEYLG 1157

RESULT 9
ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2607875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSP; P11362; IFGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
```

FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	552	565	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	141	183	ELUSGGVYIEKNDKLCMDTIDWRDIDVRDRAEIVVKNDR
FT				VC -> GQFPMVPSGLTTPQADWYLLDLDPLLTLSASSK
FT				VPVTLAAV (IN SHORT ISOFORM).
FT	VARSPLIC	184	1342	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	560	560	E -> G (IN REF. 2).
FT	CONFLICT	1064	1064	E -> G (IN REF. 2).
SQ	SEQUENCE	1342 AA;	148097 MW;	7201E7F66CA374BD CRC64;
Query Match				
Best Local Similarity 40.7%; Pred. No. 1.8e-122;				
Matches 534; Conservative 191; Mismatches 457; Indels 129; Gaps 32;				
QY	10	GLLLALLPGCAA--STQVCTGTDMKLRPLASPTHLDMLRLHYQGVQVQGNLELYLPT	67	
DB	11	GLLFLSARGEVSNSOAVCPGLTLNGLSVTGDENQYQTLKYLERCEVVMGNLEIVLTGH	70	
QY	68	NASLSFLQDIQEOGVYVLIHAHNOVROVPLQRLIRVRGTOLFEDONYALAVLDNGDPLNNTT	127	
DB	71	NADLSFLOWIREVTGVYVIVAMNEFTLPLNLRVVRGTQYDQKPAIFVW----	125	
QY	128	PVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIPHKNNQALTLTIDTN	187	
DB	126	---NSSHALRQLRLTQLTEILSGGVYIEKNDKLCMDTIDWRDIDVRDRD--AEIVVKD	178	
QY	188	RSBACHPCSPMKGSCWGESSEDDCSLRTITVCAGGC-ARCKGPLPTDCHEQCAAGCTG	246	
DB	179	NGRSCPPCHEVCKG-RCWPGSDDCQTLTKTICAPQCNGCHCFGNPNQCHDECAGCGSG	237	
QY	247	PKHSDCLACLFHNSHICELHCPALVTYNTDTPESMPNPEGRYTFGASCVCATCPYNYLST	306	
DB	238	PQDTCDFACRHNDSACVPRCPQPLVYNNKLTFLQEPNPHTKYQGVGVASCAPHNFV-V	296	
QY	307	DVGSCTVLPLHNQEVTAEDGTQRCBKSPCARVCYGLGMEHLREVRAVTSANIOEFAG	366	
DB	297	DQTSVCVRACPPDRMEVD-KNGLKWCPECGCLCPKACEGTGSG--SRFQVTDSSNIDGFVN	353	
QY	367	CKKIFOSLAFLPESFGDPSANTAPLOPOLQVFTLEITGYLTVISAWPDSLDLSVQ	426	
DB	354	CTKILGNLDFLITGLNGDPEWHKIPALDPEKLNVTREITGYINTQSWPPHMHNFVSF	413	
QY	427	NLQVIRGRILHNGAYS-LTLQIGIGISWGLRSRLRELGSGLAIHNTHTCFVHTVPDQL	485	
DB	414	NLTTIGRSLYNGRGSLLIMKNLNVTSLFRSLKUEISAGRIYISANROLCYHTHSLNWTKV	473	

QY	486	FRNPQALLHTA-NRPEDCEVGEGLACHOLCARGHCWGEGPTQCVCNCSQFLRGQECVBE	544	
DB	474	LRGPTTEERLDIKHNRPRDCVAEGKVCDCPLCSGGCGGPGQCLSCRNYSGGVCTHC	533	
QY	545	RVLQGLPREYVVARHCLPCHPECQONGSVTCFGEADOCVACAHKDPFCVCARPCSGV	604	
DB	534	NFLNGEPREFAHEAECFSCHEPCQWEGTATCNGSGSDTCAOCAHPRDCHPCVSSCPHG	593	
QY	605	KPDLSPYMPKFPDBEGACQPCFINCTHSCVDLDDKGCPEQRA-----SPLTSIVSAV	660	
DB	594	LG--AKGPIYKYPDVQNECRPCHECTQCKGPELQDCLGQTLVLIGKTHLTWALTVIAG	651	
QY	661	ILLVVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNQAOQRIKETE	719	
DB	652	--LVVIFMMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIKETE	708	
QY	720	LRKVKVLGSGFAGTVYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEAYVMVAGVS	779	
DB	709	LRKLVKLGSGVFTVHKWVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDMMLAIGSLDH	768	
QY	780	PYVSRILGICLTSTVQLVTQMLPYGCLDHDVRENRLGSGQDLNMCQIAKGMSTYLEDV	839	
DB	769	AHIVRLGLCGSSQLVQVPLPLGSLLDHVRHQRGALGPQLLNWGVQIAKGMYLEEH	828	
QY	840	RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKMALESILRRFP	899	
DB	829	GMVHRNLARNVLLKSPSQVADFGVADLLPDDKQLLYSEAKTPIKMALESIHFGKY	888	
QY	900	THQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPOPPICITDIVYTMVKW	959	
DB	889	THQSDVMSYGVTVWELMTFGAEFYAGLRLAEVDPDLLEKGERLAQOICITDIVYTMVKW	948	
QY	960	MIDSCRRFRRELSEFSRMDRDPORFVVIQNEEDLGA---SPLDSTFYRSLLEDDMDGD	1016	
DB	949	MIDENIRPTFKELANEFTEMARDPPRYLVIKES-GPGIAPGEPHGLTNKLEVELESP	1007	
QY	1017	LVDABEYLVPQOGFFCPDPAPGAGGMVHRHRSSTRSGGDLTTLGLEP-SEEEAPRSP	1075	
DB	1008	ELDLDLLEAEED-----NLATTILGSALSPLVGTNLNPRGQSLL	1048	
QY	1076	ABSEGAGSDVFDGLGMGAAGLQSLPTH-D-PSPLQRYSEDTVPPL-----SETDGYV	1128	
DB	1049	SPSSGY-MPMNGNIGESQESAVSGSSRCPRPVSLH-----PMRGLCLASESEGHV	1101	
QY	1129	A-----PLTCSQPE-----VYNQDPVRPQPPSPREGP-----L	1157	
DB	1102	TGSEAELOEKVSMCRSRSRSPRGDSAYHSQRHSLLTPTVPLSPPGLEEDVNGYVM	1161	
QY	1158	PAARFAGATLERAKTLSP-GKNGV-----KDVFAFGAVENPYLTPQGGAAPOPHPP	1210	
DB	1162	POTHLKGTSPSSREGTSSVGLSSVLGTEBED-----EYEMNRRRRHSP-PHPP	1212	
QY	1211	PAFSAFDNLYWD-----QDPPERGAPSTFKGTPTAENPEYL	1249	
DB	1213	RPSLEELGEYMDVGSGLSASLGSTQSCPLHPVIMPATAGTTPDEDYEYM	1263	
RESULT 10				
DB	ERB3_RAT	STANDARD;	PRT;	1339 AA.
AC	Q62799; Q62955;			
DT	15-DEC-1998 (Rel. 37, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)			
DE	(c-erbB3).			
GN	ERB3.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
NCBI	TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 CDNA and characterization of the
 recombinant protein.";
 RL Gene 165:279-284 (1995).
 RN [2]
 RP REVISIONS TO 85; 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroregulins and their putative receptors, ErbB2 and
 ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659 (1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U29339; AAC28498.2; --
 CC EMBL; U25230; AAC33050.1; --
 CC HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; TykC; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDF1E CRC64;
 Query Match 34.8%; Score 2369.5; DB 1; Length 1339;
 Best Local Similarity 41.0%; Pred. No. 1.1e-118;
 Matches 526; Conservative 171; Mismatches 431; Indels 155; Gaps 34;
 QY 3 LAALCRWGLLLALLPEGAA---STVCTCTDKMLRLPASPETHLDMLRLHYOGCQVQGN 59
 DB 7 LQVLC-----FLSLARGSEMGNQAVCPGTNGLSVTGDADNQYQTYLYKLYKEVVMGN 62
 QY 60 LELTLPTNLSLFLQDIOEQVGYVLIHINQVQVPLQRLIRIVRGTLQFEDNYALAVLDN 119
 DB 63 LEIVLTGHNADLSFLQWIREVTGYVLIHINQVQVPLQRLIRIVRGTLQFEDNYALAVLDN 120
 QY 120 GDPLNTTPTVGTASPGGLRELQRLSLEILKGGVLIQPNQLCYQDTILWKDIFHKNQL 179
 DB 121 ---LNYNT---NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCMDTDIMRDIRVR--- 170
 QY 180 ALTLIDTNRSRACHPCSPMKGSRGSESDCOSLTRVTVGAGGC-ARCKGPLPTDCCHE 238
 DB 171 GAEIVVKNNGANCPCHVEVCKG-RCWGPDPDQQLTKITCAPQCNGRCFGPNPNCCHD 229
 QY 239 QCAAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
 DB 230 ECAGGCGSGPQDTCFACRFNDSGACVPRCPPLVYVKNLTFTQLEPNPHTKYQYGGVCVAS 289
 QY 299 CPYNYLSTDVGSCTLVCPHNOEVTAEQTCCEKSKPKCARVCYGLGMEHLREVRVTS 358
 DB 290 CPNHFV-VDTQFCVRACPPDKMEVD-KHGLKMKCEPCGGLCPKACEGTSGS--SRYQTVD 345
 QY 359 ANIQFAGCKKIFGSLAFLESFDPGDPASNTAPLOEQLOVFEETLEETGYLYISAWPDS 418
 DB 346 SNIDGFVNCKILGNLDFLTITGLNVDPMHKKIPALDPEKUNVFRVREITGYLNIQSWPH 405
 QY 419 LPDLSVFQNLQVIRGIRLHNGAYS-LTLOGLGISWLGRLSLRELGSGLALIHNTLFCV 477
 DB 406 MNFVSFSLNTTIGRSLYNRGFSLLIMKNLVNTSLGFSLSKEISAGRVVYSANQQLCYH 465
 QY 478 HTVPWDQLPRNPHQALLHTA-NRPEDECVGEGGLACHQLCARGHCWGPGTQCNCQPLR 536
 DB 466 HSLNWTLLRGESEERLDIKYDRPLGECIAEGKVCPLDPLCSSGGCGWPGGQCLSCRNYR 525
 QY 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPC 596
 DB 526 EGVVTHCNFLQGEPEFVHEAQCFSCHEPLMGTSTCNCGSGSDACARCAHFRDGPCH 585
 QY 597 VARCPSGVKPDLSPYMPKFPDEEGACQPCINCTHSC--VDLDDKGCPEQRASPLTSI 654

growth factor receptor." ;
Nature 314:178-180(1985).
[8]
RL SEQUENCE OF 1133-1137; 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION
RN ANALYSIS.
RP MEDLINE=92038942; PubMed=1936959;
RX MEDLINE=92038942; PubMed=1936959;
RA Raz E., Schejter E.D., Shilo B.Z.;
RT "interallelic complementation among DER/fib alleles: implications for
RL the mechanism of signal transduction by receptor-tyrosine kinases." ;
RN Genetics 129:191-201(1991).
RP [9]
RX REVIEW
RN MEDLINE=97248481; PubMed=9094709;
RX Perrimon N., Perkins L.A.;
RA "There must be 50 ways to rule the signal: the case of the Drosophila
RT EGF receptor." ;
RN Cell 89:13-16(1997).
CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,
CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAP-
CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE
CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL
CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE
CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA
CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE
CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF
CC CUTICLE.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I
CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE
CC PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND
CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: UBICUOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,
CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF
CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST
CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH
CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN
CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX
CC AND THORACIC AND ABDOMINAL GANGLIA.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF052754; AAC08536.1; -
DR EMBL; AF052753; AAC08536.1; JOINED.
DR EMBL; AF052754; AAC08535.1; -
DR EMBL; AF052752; AAC08535.1; JOINED.
DR EMBL; K03054; AAA51462.1; -
DR EMBL; K03417; AAA51460.1; -
DR EMBL; K03416; AAA50965.1; -
DR EMBL; K03418; AAA51461.1; -
DR EMBL; AF109077; AAD26134.1; -
DR EMBL; AF109078; AAD26132.1; -
DR EMBL; AF109082; AAD26132.1; JOINED.
DR EMBL; AF109078; AAD26133.1; -
DR EMBL; AF109084; AAD26133.1; JOINED.
DR EMBL; AF109079; AAD26130.1; -
DR EMBL; AF109081; AAD26130.1; JOINED.
DR EMBL; AF109079; AAD26131.1; -
DR EMBL; AF109083; AAD26131.1; JOINED.
DR EMBL; AF109080; AAD26135.1; -
DR EMBL; AE003454; AAF46732.1; -
DR EMBL; K02293; CAA26157.1; -
DR EMBL; X78920; CAA55523.1; -
DR EMBL; X78918; CAA55521.1; -

[illegible]


```
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 946 ICTIDVYIMVKCWMIDSECRPRELVSFSESWARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCWMIDSECRPRELVSFSESWARDPQRFVVIQ-QGDERMHLPSPTDSKF 417
Qy 1005 YRSILEDMDGDLVDAEYILVPOQGFCDPAPGAGGWMHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYILVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSP-----APSEGAGSDVDFDGLGMAAGKLOSLPTHDPSPLOQYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHVPREDSFVQYSSDPTGN 495
Qy 1120 LPSET--DGIVAPLTCSPQPEYUNQDVRPQPSREGPLPAARPAGATLERAKTILSPGK 1177
Db 496 FLEESIDDDGFL-----PAPEYVNO--LMPKKFS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFONLYY 1222
Db 527 NQIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPSTFKGTPTAENPEYLGIDVP 1254
Db 579 WIOGNGHQLNDPNFYQDQFLPNETKPNGLLKVPAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H;
RC MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; LFQK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR Transfaser; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.2e-83;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKDPPECVACPSGVKFDLSYPIWKFPDEEGACQPCINCHSCVDLDDKKGPAEQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL--VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV--GILLVVLGVVFGILIKERQKIRKYTWRRLLQETELVEPLTPSGA 705
Db 59 NSKTPSIAAGVVGGLLCLVVGGLGGLYLR--HIVKRTLRLLQERELVEPLTPSGE 117
Qy 706 MPNQAMRILKTELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAKVLRENTSPANK 765
Db 118 APNQALRLKTEFEKKVKVLGSGAPGVYKGIWIPDGENVKIPVAKELREATSPANK 177
Qy 766 EILDEAYNAGVGSVVSRLGLICLTSTVOLYTMPLPYCGLLDHVRNGLSGSOLLNW 825
Db 178 EILDEAYNAGVGSVVSRLGLICLTSTVOLYTMPLPYCGLLDHVRNGLSGSOLLNW 237
Qy 826 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGKVP 885
Db 238 CVQIAGMNYLERRLLVHRDLAARNVLVKTPOHKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILRRFTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 946 ICTIDVYIMVKCWMIDSECRPRELVSFSESWARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCWMIDSECRPRELVSFSESWARDPQRFVVIQ-QGDERMHLPSPTDSKF 417
Qy 1005 YRSILEDMDGDLVDAEYILVPOQGFCDPAPGAGGWMHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYILVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSP-----APSEGAGSDVDFDGLGMAAGKLOSLPTHDPSPLOQYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATSNNSATNCID-----RNGQGHVPREDSFVQYSSDPTGN 495
Qy 1120 LPSET--DGIVAPLTCSPQPEYUNQDVRPQPSREGPLPAARPAGATLERAKTILSPGK 1177
Db 496 FLEESIDDDGFL-----PAPEYVNO--LMPKKPSTAM----- 524
Qy 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFD 1218
Db 525 --VQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
Qy 1219 NLYYWDQDPPPERGAPSTFKGTPTAENPEY 1248
```



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20386; AAA48760.1; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.8%; Score 1624; DB 1; Length 703;
Best Local Similarity 44.7%; Pred No. 2.2e-79;
Matches 316; Conservative 113; Mismatches 250; Indels 28; Gaps 12;

QY 8 RWGLLALLPPGAA-----STQVCTGTDMLKRLPASPETHLDMLRHLXQGVQVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKKVCCQTNNKLTQLGHVEDHFTSLQRMVYNCEVLSNLE 72

QY 62 LTYLPTNASLFLQDIOEVGYVLIHANQVRQVPLQRLRIVRGTQLFEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGYVLIANMVDVTFLENLQIRGNVLYDNSFALAVLSNH 132

QY 122 PLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDI FHKNNQAL 181
DB 133 -WNKTO-----GLRELPMKRLSELNGVGKISNNPKLCNMMDTVLWMDIITSRK-PL 182

QY 182 -TLID-TNRSRAPHCSFPMCKGSRGWESSEDQSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
DB 182 -TLID-TNRSRAPHCSFPMCKGSRGWESSEDQSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
```

```
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGEQNQCOTLTKVICAQCSGCRGKVPKSDCHNQ 242
QY 240 CAAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGYTFGASCVTAC 239
DB 243 CAAGCTGPRESDCLACRKFDDATCDDTLPPLVLYNFTTYQMDVNPBGKYSFGATCVREC 302
QY 300 PYNLSTDVGSCTLVCPHLNQEVATBDGTQRCBKCSKPCARVCYGLGMEHLREVAVTSA 359
DB 303 PHNVVTDHGSCVRSNTDTYEV-ENGVRKCKCDGLCSKVCNGIGIGELKGILINAT 361
QY 360 NIOEFAGCKKIFGSLAFPESDGDPASNTAPLOPEQLQVFFETLEITGYLYISAWPDSL 419
DB 362 NIDSFKNCTKINGDVSIILPVAFLGDAFTKTLPLDPKKLDVFTVKEISGFLLIQAWPDNA 421
QY 420 PDLISVFONLOVIRGRILHNGAYSLTLQGLGISWLGRLSRLSGLSGLALIHNNTHLCFVHT 479
DB 422 TDLYAFENLEIIRGRTKQHQYSLAVVNLKIOSLGLRSLSKESISDGDIAIMKNKNLCYADT 481
QY 480 VPWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGOE 539
DB 482 MNWRSLFATQSQKTKIIQNRKNNDCTADRHVCDPLGSDVGCWGPFGFHCSCFFSRQKE 541
QY 540 CVEECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGEADQCACAHYKDPFPC 596
DB 542 CVKQCNIQGEPEFERDSKLPCHSECLVQNSYNTTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VARCPGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCAPASQASPLTSIVS 656
DB 602 VRACPAVLGENDTL-WVKYADANAVCQLCHPNCTRGCKGPGLEGCP---NGSKTFSIAA 657
QY 657 AVV-GILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTP 702
DB 658 GVVGULLLVVVLGVVGLGILYLRRL-HIVRKRTLRRLQLQERELVEPLTP 703

Search completed: July 22, 2003, 08:06:37
Job time : 44.0519 secs
```

Qy	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSQPPEYV	1140
Db	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSQPPEYV	1140
Qy	1141	NQPDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVREPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:08:14
Job time : 97.3788 secs

THIS PAGE BLANK (USPTO)

QY 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDRV 840
Db 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDRV 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESLRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESLRRFT 900
QY 901 HQSDVMSYGYVWELMTFGAKPDGIPAREIPDLEKGERLPQPPICITIDVYIMVKCWM 960
Db 901 HQSDVMSYGYVWELMTFGAKPDGIPAREIPDLEKGERLPQPPICITIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EYLVPOQGFCCPDAPGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCCPDAPGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMGAAGLQSLTPHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMGAAGLQSLTPHDSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPERGAPPSTFKGTTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAPSPADNLYYWDQDPPERGAPPSTFKGTTAENPEYLGLDVVP 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AC
XX XX
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
DE Human
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX
XX WPI: 2002-280741/32.
XX
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
PS
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 99.9%; Score 6806; DB 23; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MELAALCRGLLLALLPPGNAASQVCTGTMKRLPASPTHLDMLRHLYQGCVQVGNL 60
Db 1 MELAALCRGLLLALLPPGNAASQVCTGTMKRLPASPTHLDMLRHLYQGCVQVGNL 60
QY 61 ELYLPTNASLSPLQDIQEVQGYVLIARHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSPLQDIQEVQGYVLIARHNRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHNKQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHNKQLA 180
QY 181 LTLIDTNRGRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRGRACHPCSPMCKGSRGWESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQDQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQDQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKIKFSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
Db 361 IQEFAGCKIKFSLAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWLGRLSRLSGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWLGRLSRLSGLALIHNTLHLCFVHTV 480
QY 481 PMDQFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Db 481 PMDQFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVYNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSPMIPKFPDEEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPMIPKFPDEEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMLIKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMLIKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPPKANKIILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSPPKANKIILDEAYVWAGVSP 780
QY 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDRV 840
Db 781 YVSRLLGICLTSTVQLTQMPYGLLDHVRNRLGSGDGLLNCWQIAKMSYLEDRV 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESLRRFT 900

Db 1021 EBYLVPOQGFCDPAPAGAGVHRRHSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080

Qy 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140

Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140

Qy 1141 NQPDVPRQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Db 1141 NQPDVPRQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200

Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPSPFTKGTPTAENPEYLGIDVVP 1255

Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPSPFTKGTPTAENPEYLGIDVVP 1255

RESULT 13

AA51143

ID AA51143 standard; Protein; 1255 AA.

XX AC AA51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KM tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key

FT Domain 1..653

FT Domain /note= "extracellular domain"

FT Domain 676..1255

FT Domain /note= "intracellular domain"

FT Domain 990..1255

FT Domain /note= "phosphorylation domain"

PN WO200212341-A2.

XX PD 14-FEB-2002.

XX PF 03-AUG-2001; 2001WO-US24283.

XX PR 03-AUG-2000; 2000US-0632507.

XX PA (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cheever MA, Gheysen D;

XX DR WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

PT or enhancing an immune response to the protein, has Her-2/neu

PT extracellular domain fused to Her-2/neu intracellular or

PT phosphorylation domain

XX PS Claim 68; Fig 7; 14lpp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

CC or c-erbB2), an oncogenic self-protein and target for anti-cancer

CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

CC in a variety of cancers, including breast, ovarian, colon, lung and

CC prostate cancer. Her-2/neu is a member of the tyrosine kinase

CC family of receptor-like glycoproteins. It comprises an extracellular

CC domain with homology to the epidermal growth factor receptor

CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

CC intracellular domain that also shows homology to EGFR. Its

CC overexpression correlates with a poor prognosis in breast and

CC ovarian cancers. The invention provides Her-2/neu fusion

CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In

CC preferred fusion proteins, the extracellular domain of a Her-2/neu

CC protein is fused to a Her-2/neu intracellular domain or

CC phosphorylation domain (or its DeltapD fragment). An immune

CC response to Her-2/neu protein is elicited or enhanced by

CC administering the fusion protein in the form of a vaccine, or by

CC transfected cells of an animal ex vivo with a nucleic acid

CC encoding the fusion protein, and delivering the transfected cells

CC to the animal. The fusion proteins, nucleic acids, and isolated

CC specific T-cells are useful for inhibiting the development of a

CC cancer, especially breast, ovarian, colon, lung or prostate cancer

CC in a patient. T cells that specifically react with a Her-2/neu

CC fusion protein can be used to remove tumour cells from a sample in

CC order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 23; Length 1255;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLAPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60

Db 1 MELAALCRWGLLLALLAPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHNKOLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHNKOLA 180

Qy 181 LTLIDNTRSRACHPCSPMCKGRCWGESSEDCOSLRTVCAGCARCKGLPDTDCHEQC 240

Db 181 LTLIDNTRSRACHPCSPMCKGRCWGESSEDCOSLRTVCAGCARCKGLPDTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 IQEFAGCKITFGSLAFLPESFDGPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420

Db 361 IQEFAGCKITFGSLAFLPESFDGPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTLHCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWGLSLRELGSGLALIHNTLHCFVHTV 480

Qy 481 PWDQFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Db 481 PWDQFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600

Db 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600

Qy 601 PSQVKPDLSPYMPWKFPDEBEGACQPCINCTHSCVDLDDKGCQPAEQASPLTSISAVVG 660

Db 601 PSQVKPDLSPYMPWKFPDEBEGACQPCINCTHSCVDLDDKGCQPAEQASPLTSISAVVG 660

Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLRLQBELVEPLTPSGAMPNQAMRILKETEL 720

Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLRLQBELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRNTSPKANKETLDEAYVMAGVSP 780

Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRNTSPKANKETLDEAYVMAGVSP 780

Db 1141 NQEDVRQPSREGPLPAARPGATLRLERKTLSPGKNGVVKDVAFGGAVENPEYLTTPQ 1200
QY 1201 GGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAFSPAFDNLYYWDQDPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
ID AAE20479 standard; Protein; 1255 AA.
XX AAE20479;
AC AAE20479;
XX 01-JUL-2002 (first entry)
XX Human Her-2/neu protein.
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX WO200214503-A2.
XX 21-FEB-2002.
XX 14-AUG-2001; 2001WO-US41733.
XX 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX (CORI-) CORIXA CORP.
PA Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
PS Disclosure; Page 114-117; 129pp; English.
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions, for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 23; Length 1255;

Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MELAAALCRGGLLALALPPGAASQVCTGDMKRLRPASPTHLDMLRHLHYOGQVVOGNL 60
Db 1 MELAAALCRGGLLALALPPGAASQVCTGDMKRLRPASPTHLDMLRHLHYOGQVVOGNL 60
QY 61 ELTYLFTNASLSFLQDIEQVQGVVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLFTNASLSFLQDIEQVQGVVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCOSLRTVCAGGCARCKGPLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGRCWGESSEDCOSLRTVCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACILHFNHSGICEILHCPALVTYNTDTFESMENPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACILHFNHSGICEILHCPALVTYNTDTFESMENPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOQVFPETLEEITGVLYISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOQVFPETLEEITGVLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSLRELASGLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSLRELASGLALIHNTLHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKPPFCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKPPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDILLNWCMLAKGMSYLEDVR 840
Db 781 YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDILLNWCMLAKGMSYLEDVR 840
QY 841 LVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALESILRRRFT 900
Db 841 LVHRDLAARNVLVSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALESILRRRFT 900
QY 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPCTTIDVYMWKCM 960
Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPCTTIDVYMWKCM 960
QY 961 IDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFFVRSLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSEFMRARDPQRFVVIQNEIDLGPASPLDSTFFVRSLEDDMDGLVDA 1020
QY 1021 EBYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDI/TLGLEPSEEPAPSLPSEG 1080

Db 1141 NQDVRQPSRSGPLPAARPACATLERPKTSLSPKNGVVKVAFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW tumour response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN W0200141787-A1.
PD 14-JUN-2001.
XX
XX 11-DEC-2000; 2000WO-US33591.
XX
XX 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
XX WPI; 2001-374995/39.
XX
XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
XX Disclosure; Page 15; 199pp; English.
XX
XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;

Query Match 99.7%; Score 6698; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPEHLDMLRHLHLYQGQVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPEHLDMLRHLHLYQGQVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGKPLQVIVKANSK 240
Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGKPLQVIVKANSK 240

Qy 241 FIGITELKHSDDLACILHFNHSGICELHCPALVYNTDTTFESMPNPBGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDDLACILHFNHSGICELHCPALVYNTDTTFESMPNPBGRTYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360

Qy 361 IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVPELTLEETIGYLYISAWPDSLP 420
Db 361 IOEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVPELTLEETIGYLYISAWPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWGLSLRELGLSLALIHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGTSWGLSLRELGLSLALIHNTHLCFVHTV 480

Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSOFLRQEC 540
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCSOFLRQEC 540

Qy 541 VEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600
Db 541 VEECRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600

Qy 601 PSGVKPDLSTYMPITWKFPEDEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTYMPITWKFPEDEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660

Qy 661 ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOTELVLEPLTPSGAMPNQOMRILKETEL 720
Db 661 ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOTELVLEPLTPSGAMPNQOMRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMAGVGP 780

Qy 781 YVSRLLGICLTSTVQLVTQTLMPYGCILLDHVRENRLGSGDILLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQTLMPYGCILLDHVRENRLGSGDILLNWCQIAKGSYLEDVR 840

Qy 841 LVHRDLAARNVLKSPNNHVKITDIFGLARLLDDETHADGGKVPKKNMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNNHVKITDIFGLARLLDDETHADGGKVPKKNMALESILRRRFT 900

Qy 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYMIWKVCM 960
Db 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYMIWKVCM 960

Qy 961 IDSECEPRFRELVSFSESRMARDPQRFVWIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECEPRFRELVSFSESRMARDPQRFVWIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020

Qy 1021 EYLVPEQQGFCCDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEBAPSLAPSEG 1080
Db 1021 EYLVPEQQGFCCDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEBAPSLAPSEG 1080

Db 1021 EYLVPQQFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLQSLPTHDPQLORYSEDPVLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPTHDPQLORYSEDPVLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQDVRQPSPRSGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Db 1141 NQDVRQPSPRSGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPYLGLDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPERGAPSTFKGTPTAENPYLGLDVPV 1255

RESULT 11

AAE24067
ID . AAE24067 standard; Protein; 1255 AA.

AC AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

XX WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

XX Bennett CF, Cowsett LM;

XX WPI; 2002-471192/50.

XX N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.

XX Sequence 1255 AA;

Qy Query Match 98.7%; Score 6698; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKILRLPASPTHLDMLRHLVGGCCVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKILRLPASPTHLDMLRHLVGGCCVQGNL 60

Qy 61 ELTYLPTNASLFLQDIOEQVQVYLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIOEQVQVYLIAHNQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDILFKKNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDILFKKNOLA 180
Qy 181 LTLIDTNRSPACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRSPACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLPOYIKANSK 240
Qy 241 FIGITELKESDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKIFGLSLAFLPESFDGDPASNTAPLOEQVFEETLEETGYLIYSAMPDLSL 420
Db 361 IOEFAGCKIFGLSLAFLPESFDGDPASNTAPLOEQVFEETLEETGYLIYSAMPDLSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWGLSLRELGLSLALIHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGTSWGLSLRELGLSLALIHNTHLCFVHTV 480
Qy 481 PWDOLFNRPHQALLHTANRPEDECVEGLACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Db 481 PWDOLFNRPHQALLHTANRPEDECVEGLACHOLCARGHCWGPGPTQCVNCSOFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVCAHYKDPFCVARC 600
Qy 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQSTELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQSTELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNLYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNLYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
Qy 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
Qy 1021 EYLVFPQQFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db 1021 EYLVFPQQFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLQSLPTHDPQLORYSEDPVLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPTHDPQLORYSEDPVLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQDVRQPSPRSGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO 1200

Db 1141 NQPDVRRPSPREGPLPAARPAATLRLPKTLSPGKGVVVKDVFARGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX
PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
XX WPI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
PT
XX Disclosure; Page 114-117; 129pp; English.
XX
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match

98.7%; Score 6698; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MELAALCRWGLIALALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
Db 1 MELAALCRWGLIALALLPPGAASQVCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGYVLIHQNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIHQNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODILKWDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODILKWDIFHKNNOLA 180
QY 181 LTLIDTNRGRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPCKSGRCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHEQC 240
QY 241 FIGITELKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKIFGSLAFPLPESFGDGPASNTAPLQPEOLQVFETLEETITGYLYISAWPDSL 420
Db 361 IQEFAGCKIFGSLAFPLPESFGDGPASNTAPLQPEOLQVFETLEETITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGISMGLSLRELGSGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGISMGLSLRELGSGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSFLRGQEC 540
QY 541 VEECRVLOGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVVARC 600
Db 541 VEECRVLOGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVVARC 600
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPASPLTSIVSAVVG 660
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAGFVYVKGWIPOGENVKIPVAIKVLRNTSPKANKIIDEAYVMAGVGP 780
Db 721 RKVKVLGSGAGFVYVKGWIPOGENVKIPVAIKVLRNTSPKANKIIDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRNRRGLSQDILNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRNRRGLSQDILNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKKNMALESILRRRFT 900
QY 901 HQSDVMSYGVVWELMTFCAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVVWELMTFCAKPYDGIIPAREIPDLLEKGERLPOPPCTIDVYIMVKCWM 960
QY 961 IDSECRPRFELVSEFSRMDPQRFVWJQNEDLGPASPLDSTFYRSLDEDDMGDLVDA 1020
Db 961 IDSECRPRFELVSEFSRMDPQRFVWJQNEDLGPASPLDSTFYRSLDEDDMGDLVDA 1020
QY 1021 BEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG 1080
Db 1021 BEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG 1080

Db 1021 EEVLVPOQGFCDPAPACAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 Qy 1081 AGSDVFDGLGMAKGLQSLPHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAKGLQSLPHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKOVFAFGGAVENPEYLTPO 1200
 Db 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKOVFAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPYLGLDVPV 1255
 Db 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPYLGLDVPV 1255

RESULT 13

AAMS1143
 ID AAMS1143 standard; Protein; 1255 AA.

XX AC AAMS1143;
 XX 17-JUN-2002 (first entry)
 DT Human Her-2/neu oncogene-encoded p185 glycoprotein.
 DE Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 KW tyrosine kinase; receptor; c-erbB2; gene therapy.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Domain 1..653
 FT /note= "extracellular domain"
 FT Domain 676..1255
 FT /note= "intracellular domain"
 FT Domain 990..1255
 FT /note= "phosphorylation domain"

XX W0200212341-A2.
 XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.
 XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Cheever MA, Gheysen D;
 DR WPI; 2002-241743/29.
 DR N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
 or enhancing an immune response to the protein, has Her-2/neu
 extracellular domain fused to Her-2/neu intracellular or
 phosphorylation domain
 Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
 or c-erbB2), an oncogenic self-protein and target for anti-cancer
 vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 in a variety of cancers, including breast, ovarian, colon, lung and
 prostate cancer. Her-2/neu is a member of the tyrosine kinase
 family of receptor-like glycoproteins. It comprises an extracellular
 domain with homology to the epidermal growth factor receptor
 (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 intracellular domain that also shows homology to EGFR. Its
 overexpression correlates with a poor prognosis in breast and
 ovarian cancers. The invention provides Her-2/neu fusion
 proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
 preferred fusion proteins, the extracellular domain of a Her-2/neu
 protein is fused to a Her-2/neu intracellular domain or
 phosphorylation domain (or its DeltaPD fragment). An immune
 response to Her-2/neu protein is elicited or enhanced by
 administering the fusion protein in the form of a vaccine, or by
 transfecting cells of an animal ex vivo with a nucleic acid
 encoding the fusion protein, and delivering the transfected cells
 to the animal. The fusion proteins, nucleic acids, and isolated
 specific T-cells are useful for inhibiting the development of a
 cancer, especially breast, ovarian, colon, lung or prostate cancer
 in a patient. T cells that specifically react with a Her-2/neu
 fusion protein can be used to remove tumour cells from a sample in
 order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHQNQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARGCKGPLQVIRANSK 240
 Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARGCKGPLQVIRANSK 240
 Qy 241 FIGITELKHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IQEFAGCKIFGSLAFPLPSFDGDPASNTAPLOEQVETLEEITGYLIISAWPDSL 420
 Db 361 IQEFAGCKIFGSLAFPLPSFDGDPASNTAPLOEQVETLEEITGYLIISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNHQALLHTANRPDECVGEGGLACHOLCARGHCWGPGPTQCVNCSFLRQEC 540
 Db 481 PWDQLFRNHQALLHTANRPDECVGEGGLACHOLCARGHCWGPGPTQCVNCSFLRQEC 540
 Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFEGPEADQCVACHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFEGPEADQCVACHYKDPFCVARC 600
 Qy 601 PSGVKPDLSVMPITWKEPDEEGACQPCINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660
 Db 601 PSGVKPDLSVMPITWKEPDEEGACQPCINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780

QY 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVMIMVKCWM 960
QY 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
Db 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVDA 1020
QY 1021 EYLVPQOQFFCDDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
Db 1021 EYLVPQOQFFCDDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLTPO 1200
Db 1141 NQPDVRRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVAFPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
FN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR N-PSDB; ABK10730.
DR
DR
XX
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX
PS Disclosure; Page 71-74; 74pp; English.
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQVGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVQVGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVYLIHQAQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIHQAQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLREQLRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTTPTVGASPGGLREQLRLSLEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCQSLTRTVCCAGCCKPQVIRKANSK 240
Db 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCQSLTRTVCCAGCCKPQVIRKANSK 240
QY 241 FIGITELKHSDDLCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPLOPEOLQVFETLEETIGVLYISAWPDSL 420
Db 361 IOEFAGCKKI FGSIAFLPESFDGDPASNTAPLOPEOLQVFETLEETIGVLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILLHNGAYSILTLQGLGSIWGLSLRLSRELGSGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILLHNGAYSILTLQGLGSIWGLSLRLSRELGSGLALIHNNHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVARC 600
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVARC 600
QY 601 PSGVKPDLSTYMPITWKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLSTYMPITWKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYMAGVGP 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELIDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 Qy 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Qy 1021 EYLVPQOQFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Db 1021 EYLVPQOQFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVQDVAFGGAVENPEYLTPO 1200
 Db 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVQDVAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255
 Db 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 15

AAR39568
 ID AAR39568 standard; Protein; 1433 AA.

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX WO9316185-A.

XX 19-AUG-1993.

PF 05-FEB-1993; 93WO-US01055.

XX 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.

PA (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

XX N-PSDB; AAQ46083.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87pp; English.

CC c-erbB-2 refers to a protein antigen expressed on the surface of
 CC tumour cells, such as breast and ovarian tumour cells, which is an
 CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
 CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
 CC the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 98.0%; Score 6655; DB 14; Length 1433;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
 Qy 1 MELAALCRGLLALLAPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLYQGCQVQGNL 60
 Db 1 MELAALCRGLLALLAPPGAASQVCTGTDMKRLRLPASPTHLDMLRHLYQGCQVQGNL 60
 Qy 61 ELTYLPTNASLSFLODIQEVQGVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLODIQEVQGVLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGSLRELQLRSLTEILKGGVLIQORNPOLCYQDTILWKDIPHKKNQLA 180
 Db 121 DPLNNTPTVTGASPGSLRELQLRSLTEILKGGVLIQORNPOLCYQDTILWKDIPHKKNQLA 180
 Qy 181 LTLIDNTRACHPCSPMCKGRCWGESSEDCOSLTRVCAGGCARCKGPLQYIKANSK 240
 Db 181 LTLIDNTRACHPCSPMCKGRCWGESSEDCOSLTRVCAGGCARCKGPLQYIKANSK 240
 Qy 241 FIGITELKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVSGSCTLVCPHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IOEFAGCKTFCGLAEFLPESFDGDPASNTAPLOPELOVFPETLEEITGVLYISAWPDSL 420
 Db 361 IOEFAGCKTFCGLAEFLPESFDGDPASNTAPLOPELOVFPETLEEITGVLYISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWGLSLRLSRELGSGLALIHNTHLFCFVHT 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWGLSLRLSRELGSGLALIHNTHLFCFVHT 480
 Qy 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCOSFLRGQEC 540
 Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVNCOSFLRGQEC 540
 Qy 541 VEECRVLOGLPREYVNAHCLPCHPSCQPNQSVTCFGEADOCVACAHKDPFCVARC 600
 Db 541 VEECRVLOGLPREYVNAHCLPCHPSCQPNQSVTCFGEADOCVACAHKDPFCVARC 600
 Qy 601 PSGVKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCPEAQASPLTSISAVVG 660
 Db 601 PSGVKPDLSPYMPKPFDEEGACQPCINCTHSCVDLDDKGCPEAQASPLTSISAVVG 660
 Qy 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLOSTELVEPLTPSGAMPNOAQRILKETEL 720
 Db 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLOSTELVEPLTPSGAMPNOAQRILKETEL 720
 Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
 Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNYSILEDVR 840
 Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNYSILEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEVHADGGKVPKIMMALESILRRFT 900
 Qy 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 Db 901 HQSDVWSYGVTVWELMTFGAKPDGIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
 Qy 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMDGLVDA 1020
 Qy 1021 EYLVPQOQFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
 Db 1021 EYLVPQOQFFCPDPAPGAGGVMHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080

Qy	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	.AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVNV	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVNV	1255

Search completed: July 22, 2003, 08:40:50
 Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec
Title: SEQ4-325-339-12
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVP 1255
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : PIR 73: *
1: piri: *
2: piri: *
3: piri: *
4: piri: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				%		Query Match		Length DB ID		Description	
Result No.	Score	Match	Length	DB	ID						
1	6728	98.7	1255	1	A24571	protein-tyrosine k					
2	5931	87.0	1260	1	TVRTNU	protein-tyrosine k					
3	5921.5	86.9	1254	2	I48161	p-185 precursor -					
4	3146	46.2	1210	1	GQHUE	epidermal growth f					
5	3117	45.7	1210	2	A53183	epidermal growth f					
6	3094.5	45.4	1223	1	TVCHLV	epidermal growth f					
7	2966.5	43.5	1308	2	A47253	epidermal growth f					
8	2669	39.2	1166	1	S06142	protein-tyrosine k					
9	2414.5	35.4	1342	2	A36223	kinase-related tra					
10	2328.5	34.2	1339	2	JC4387	epidermal growth f					
11	1766.5	25.9	698	1	TVFVLV	protein-tyrosine k					
12	1703	25.0	604	1	TVYVH	protein-tyrosine k					
13	1647	24.2	544	2	S35745	protein-tyrosine k					
14	1640	24.1	545	2	S00727	kinase-related tra					
15	1630.5	23.9	1330	1	GQFFE	epidermal growth f					
16	1623	23.8	540	2	B44776	protein-tyrosine k					
17	1621	23.8	540	1	TVFVEB	protein-tyrosine k					
18	1509	22.1	644	2	A36325	epidermal growth f					
19	1290	18.9	1323	2	E88257	protein let-23 [im					
20	1290	18.9	1374	2	S70712	protein-tyrosine k					
21	1204	17.7	1369	2	S70712	protein-tyrosine k					
22	1162	17.1	1717	1	A45558	epidermal growth f					
23	1126	16.5	527	2	A42032	epidermal growth f					
24	975.5	14.3	843	2	A27131	epidermal growth f					
25	806.5	11.8	346	2	S13807	protein-tyrosine k					
26	754.5	11.1	311	2	S13808	protein-tyrosine k					
27	727	10.7	1363	2	T43220	insulin-like growt					
28	697	10.2	1382	1	INHUR	insulin receptor p					
29	696	10.2	1383	2	A36080	insulin receptor p					

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A4188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:2903577
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:K03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HRR2 (neu) promoter: evidence for multiple mechanisms for transcriptiona
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

insulin receptor p
insulin-like growt
protein-tyrosine k
insulin receptor-r
insulin receptor-r
insulin-like growt
insulin-like growt
insulin receptor -
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
protein-tyrosine k
protein-tyrosine k

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAAS8637.1; PID:G553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NGL; NEU; HER-2
 A;Cross-references: GDB:I20613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.7%; Score 6728; DB 1; Length 1255;
 Best Local Similarity 98.6%; Pred. No. 5.6e-267;
 Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy	1	MELAAACRWGLLALLPPGAASCTGCTDMKRLPASPEHLDMRLHYGCGOVVGNL	60
Db	1	MELAAACRWGLLALLPPGAASCTGCTDMKRLPASPEHLDMRLHYGCGOVVGNL	60
Qy	61	ELTYLPTNASLFLQDIQEVGYVLIAHNQVRQVPLQRLIRVGTQFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLFLQDIQEVGYVLIAHNQVRQVPLQRLIRVGTQFEDNYALAVLDNG	120
Qy	121	DPLNNTTPTVTCASPGGLREQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA	180
Db	121	DPLNNTTPTVTCASPGGLREQLRSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA	180
Qy	181	LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC	240
Db	181	LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC	240
Qy	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLCPLHNVQVTAEDGTQRCCKSPCARVCYGLGMOYKANSKFTGIT	360
Db	301	YNYLSTDVGSCTLCPLHNVQVTAEDGTQRCCKSPCARVCYGLGMOYKANSKFTGIT	360
Qy	361	ELEPAGCKKFGSLAFIPESPDGPASNTAPLQPEQLQVPELLEITGYLIVISAWPSL	420
Db	361	IQEPAGCKKIFGSLAFIPESFDGPASNTAPLQPEQLQVPELLEITGYLIVISAWPSL	420
Qy	421	DLVSFQNLQVIRGILHNGAYSLTLOGIGISWGLRLSRLGSLALIHNNTHLCFVHTV	480
Db	421	DLVSFQNLQVIRGILHNGAYSLTLOGIGISWGLRLSRLGSLALIHNNTHLCFVHTV	480
Qy	481	PWDQLFRNPQALLHTANRDECEVCGGLACHQICARGHCWGPGPTQCVNCSQPLRQEC	540
Db	481	PWDQLFRNPQALLHTANRDECEVCGGLACHQICARGHCWGPGPTQCVNCSQPLRQEC	540
Qy	541	VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGRPADQCCVCAHYKDPFPFCV	600
Db	541	VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGRPADQCCVCAHYKDPFPFCV	600
Qy	601	PSGVKPDLSYMPIWKFPDEEGACQCPINCTHSCVDLDDKCPAEQASPLTSISAVVG	660
Db			

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746

R;Masui, T.; Mann, A.M.; Macatee, T.E.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no

2-thiazolylformamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663, 'V', 665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.0%; Score 5931; DB 1; Length 1260;
Best Local Similarity 87.0%; Pred. No. 1.6e-234;
Matches 1093; Conservative 52; Mismatches 110; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 4 MELAAWCRWGLFLLALLPPGIAGTQVCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLSFLQDIQEVQVYLIHANOVRVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 64 ELTYVPAASLSFLQDIQEVQVYLIHANOVRVPLQRLRIVRGTLQFEDNYALAVLDNR 123
QY 121 DPLNNTTPTVT-GASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 179
DB 124 DPQDNVAASPTGRTPEGLRELQRLSLTEILKGGVLIQORNPOLCYQDMVLWKDIFHKNNQL 183
QY 180 ALTLIDNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGACRCKGLPTDCHEQ 239
DB 184 APVDIDNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGACRCKGLPTDCHEQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PYNLSTDVGSCTLVCPLNHQNVEYTAEDGTQRCBCKSPCARVCYGLGMQVIKANSFPGI 359
DB 304 PYNLSTEVGSCTLVCPNNQVEYTAEDGTQRCBCKSPCARVCYGLGMHLRGARITSD 363
QY 360 TELEFAGCKIKI FGSFLAPLPSFGDPSANTAPLOPEOLQVFETLEITGLYLIISAWPDSL 419
DB 364 NVQEFDPCKKIFGSLAPLPSFGDPSANTAPLOPEOLQVFETLEITGLYLIISAWPDSL 423
QY 420 PDLVSFQNLQVIRGRIHNGAYSILTLQGLGISMGLRSRELGLALIHNTLHCFVHT 479
DB 424 RDLVSFQNLRIIRGRIHNGAYSILTLQGLGISMGLRSRELGLALIHNTLHCFVHT 483
QY 480 VPMDQLFRNPQALLHTANRPEDE-CVYEGGLACHQLCARGHCWGPQTCVNCQSFLRGQ 538
DB 484 VPMDQLFRNPQALLHSGNRPEDLCVSSGLVCNLSLACHGCHWGPQTCVNCQSFLRGQ 543
QY 539 ECVEECRVLOGLPREYVNAHCLPCHPECOQNGSVTCFQPEADQCVACHYKDPDFCVA 598
DB 544 ECVEECRVWKGLPREYVSDKRCLPCHPECOQNSSETCFGSEADQCAACHYKDSSCVA 603
QY 599 RCPGSKVPDLSYMPIWKFPDEEGACOPCINCTHSCVDLDDKCPAEORASPLTSIVSAV 658
DB 604 RCPGSKVPDLSYMPIWKFPDEEGICQPCINCTHSCVDLDDERCPAEORASPLTFIATV 663
QY 659 VGILLVVGLVWFGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKET 718
DB 664 EGVLFLIILVVVGLIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKET 723
QY 719 ELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 778
DB 724 ELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 783
QY 779 SPVVSRLIGLICLTSTVOLVTQMPYGCGLLDHVENRGRLGSODLLNCKMIOIAKMSVLE 838
DB 784 SPVVSRLIGLICLTSTVOLVTQMPYGCGLLDHVENRGRLGSODLLNCKMIOIAKMSVLE 843
QY 839 VRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 898
DB 844 VRLVHRDLAARNVLKSPNPKVITDFGLARLLDIDETEHADGGKVPKIMMALESILRRR 903
QY 899 FTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWVK 958

DB 904 FTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMIWVK 963
QY 959 WMIDSECRPRFRELVESEFARMARDPQRFVVIQNEIDLGPSPLDSTFYRSILLEDDMGDLV 1018
DB 964 WMIDSECRPRFRELVESEFARMARDPQRFVVIQNEIDLGPSPLDSTFYRSILLEDDMGDLV 1023
QY 1019 DAEYLVPOQGFPCPDPAFGAGGMVHRRHSSTRSGGDLTLGLPSEEEAPRSLAPS 1078
DB 1024 DAEYLVPOQGFPCPDPTCTGTAHRRHSSTRSGGDLTLGLPSEEEAPRSLAPS 1083
QY 1079 EGAGSDVDFGDLGMAAGKGLQSLPHTDPSLORYSEDPVPLPSETDGYVAPLTCPOPE 1138
DB 1084 EGAGSDVDFGDLGMAAGKGLQSLPHTDPSLORYSEDPVPLPSETDGYVAPLTCPOPE 1143
QY 1139 YVNPQDVRPQPPSPREPGLPAARPGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLT 1198
DB 1144 YVNOSEVQPPPLTPGGLPPVPRPGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLV 1203
QY 1199 POGGAAPQHPPPAFSPAFDNLYYMDQPPERGAAPTFTKGTPTAENPEYLGLDVPV 1255
DB 1204 PREGTASPPHPSPAFSPAFDNLYYMDQSPSPSPNFEPTAENPEYLGLDVPV 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PIDN:BAA03801.1; PID:g747595
C:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.9%; Score 5921.5; DB 2; Length 1254;
Best Local Similarity 86.6%; Pred. No. 3.9e-234;
Matches 1087; Conservative 61; Mismatches 106; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAAWCRWGLLLALLPPGAASTQVCTGDMKLRLPASPETHLDIVRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQVYLIHANOVRVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYLPAANATLSFLQDIQEVQVYLIHANOVRVPLQRLRIVRGTLQFEDNYALAVLDNR 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLDNVTATATGTPTPEGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGACRCKGLPTDCHEQC 240
DB 181 PVDIDNRSRACHPCSPCKGSRGWESSEDCQSLTRTVACGACRCKGLPTDCHEQC 240
QY 241 AAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 301 YNVLSTDVGSCTLVCPLNHQNVEYTAEDGTQRCBCKSPCARVCYGLGMQVIKANSFPGI 360
DB 301 YNVLSTEVGSCTLVCPLNHQNVEYTAEDGTQRCBCKSPCARVCYGLGMHLRGARITSD 360

QY 361 ELRPAAGCKITFGSLAFIPESGDGPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420
Db 361 IOEPAGCKITFGSLAFIPESGDGNPSSGIAPLTPPEQLOVFETLEEITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLOGLGISWGLRSLRGLSLALIHNTLHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRVLDHGYSLALQGLGIRWLGRSLRGLSLALIHNTLHLCFVHTV 480
QY 481 PWDLFRNPQALLHTANRDEDECVBEGGLACHQLCARGHCWGPGPTQCVNCSFRLRQEC 540
Db 481 PWDLFRNPQALLHSGNPSEECGLKDFACYPICAHGHCWGPGPTQCVNCSFRLRQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQVCAHYKDPFFCVAR 600
Db 541 VKECRVWGLPREVYNARHCLPCHPECPQNGSVTCFGEADQVCAHYKDPFFCVAR 600
QY 601 PSQVKPDLNYPINWKFDEEGACQPCPINCCHSCVDLDDKCPAEQQRASPLTSTVSVA 660
Db 601 PSQVKPDLNYPINWKFDEEGACQPCPINCCHSCVDLDDKCPAEQQRASPLTSTVSVA 660
QY 661 ILLVVVGVLGILIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
Db 661 ILLFLVIGVVVGLIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQMRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 780
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGV 780
QY 781 YVSRLLGICLTSTVOLTPQCLLDHVRNKRGLSGDOLLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVOLTPQCLLDHVRNKRGLSGDOLLNWCQIAKGSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900
QY 901 HQSDVWSVGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWK 960
Db 901 HQSDVWSVGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWK 960
QY 961 IDSECRPRFRELVSFESEFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLED 1020
Db 961 IDSECRPRFRELVSFESEFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLED 1020
QY 1021 EYLVPQGFPPDPAPCAGVMVHRHRSSTRSGGDLTLGLPSEEEAPRSPAPSE 1080
Db 1021 EYLVPQGFPPDPAPCAGVMVHRHRSSTRSGGDLTLGLPSEEEAPRSPAPSE 1080
QY 1081 AGSDVFDGLGMGAAGKQSLPTHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQ 1140
Db 1081 AGSDVFEGLMGATGQPSISPRDLSPQRYSEDPTLPLPTEYDGVVAPLTCSPQ 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPE 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPE 1200
QY 1201 GGAAPOHPHPPAFPAFDNLYWQDPPERGAPSTFKGPTTANPEYGLGDV 1255
Db 1201 GGASQPH-PPALCPAFDNLYWQDPPERSGPPNTEGPTAENPEYGLGDV 1254

RESULT 4

GQHUE

epidermal growth factor receptor precursor - human

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Homo sapiens (man)

C;Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C;Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

Nature 309, 418-425, 1984

A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210 <ULL>
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g5757924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Shii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epidermal gr
A;Reference number: A25772; MUID:85270438; PMID:2391899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:gl81981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript termina
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:gl81977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Me
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 150-187, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-32
'798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF rece
R;Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal gro
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744, 'X', 746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superc
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:212608
 R:Heiseremann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: I49643; MUID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.7%; Score 3117; DB 2; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 6e-120;
 Matches 634; Conservative 168; Mismatches 359; Indels 112; Gaps 25;
 Qy 11 LLLALLPPGAA--STQVCTGDMKRLPASPETHDMLRHLYQCGVQVQGNLEITYLPTN 68
 Db 14 LTLALCAAGGAALEKKVCCQTSNRLTQLGTGFEDHFLSLQRMNCEVVLGNLEITYVQRN 73
 Qy 69 ASLSFLQDIQEVQGVLYIAHNOVRQVPLRLRIVRGTOLFDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLGFLKTIQEVAGVLYIALNTVERIPLNIQIRGNALYENTYALILSN----- 124
 Qy 129 VTGASPGRLQLRLSLTEILKGGVLIQRNQLCVQDTILWKDI----FHKNNALALTLI 184
 Db 125 -YGNRTGURELPNRELQELILIGAVRPSNNPILCNWDIIQRDIVQNVFMSNMDL---- 180
 Qy 185 DTNRSRACHPCSPMCKGRWCSESDCOSLTRTVACGGCA-RCKGPLPTDCHEQCAAG 243
 Db 181 -QSHPSSCPCKDPSCPNGSCNGGGBEENCKLTKIICAQCCSHRCRGRSPSCDCHNQCAAG 239
 Qy 244 CTGPKHSDCLALFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
 Db 240 CTGPRESCLVCQKQFQDEATCKOTCPPLMLNPTTYQMDVNPPEKYSFGATCVKCKPRNY 299
 Qy 304 LSTDVGSCTLCPHLNQVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFTIGITELE 363
 Db 300 VVTDHGSVCRAGPDYIEV-EDGIRCKCKCDGCRKVCNGIGIGFEK-DTILSNATNIK 357
 Qy 364 -FAGCKITFGSLAFIPESFGDGPASNTAPQEQLOVPEETLEITGYLYISAWPDSLPDL 422
 Db 358 HFKYCTAISGDLHILPFAFKGDSFTRTPPLDPRLEILKTVKEITGFLIIQAWPDNWTDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C:Species: Gallus gallus (chicken)

C:Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A00643

R:Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A:Reference number: A27720; MUID:88261272; PMID:3260329

Qy 423 SVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGSGLAIHHNTHLCFVHTVPW 482
 Db 418 HAFENLEIIRGRTKQGFSLAVVGLNITSLGRLSLAKEISDGDVITSGNRNLVCYANTINW 477
 Qy 483 DQLFRNPQALLHTANRPEDECYEGEGIACHQLCARHGCGPGTQCVCNCQFQLRGQBCVE 542
 Db 478 KKLFGTNPQTKIMNRAEKDKAVNHVCNPLCSSEGCWGPEDRDCVSCQNVSRGRCVE 537
 Qy 543 ECRVLQGLPREYVNAHCLPCHPECQDQNSVTCFGEADQCACAHYKDPFPFCVACPS 602
 Db 538 KCNILEGEPREFVENSECICHPCECLPQAMNITCTGRGPDNCICQAHYIDGPHCVKTCPA 597
 Qy 603 GVKPDLSYPMIWFPEDEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSIVSAVVGIL 662
 Db 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPLQGCCEVWSPGPKFISATGIYVGL 656
 Qy 663 LVVVLGVVFGI-LIKRRQKIRKYTRRLLEQETELVEPLTPSGAMPNQAOQMRILKETELR 721
 Db 657 LFIVV-VALGIGLFMRERHIVRKTRTLRLQLQERELVEPLTPSGEAPNQAHRLILKETEFK 715
 Qy 722 KVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGSFY 781
 Db 716 KIKVLGSGAGTGVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPH 775
 Qy 782 VSRLLGICLTSTVOLVTQLMPYGCLLDHDVRENRGLSGQDLLANWCMOIAKGMSTYLEDVRL 841
 Db 776 VCRLLGICLTSTVOLITQLMPYGCLLDYVREHNDNIGSYLLNWCVCQIAKGMNLEDRRL 835
 Qy 842 VHRDLAARNVLKSPNHNKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFRTH 901
 Db 836 VHRDLAARNVLKTPQHVKITDFGLAKLGLCAEKEKHYHAEGKVPKIKWMALESILHRIYTH 895
 Qy 902 QSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVWKCMWI 961
 Db 896 QSDVMSYGVTVWELMTFGSKPYDGPASDISILEKGERLPQPPICTIDVYIMVWKCMWI 955
 Qy 962 DSCRPREFRELSEFSMARDPORFVVIQ-NEDLGASPLDSTFYRSLLEDGDDGLVDA 1020
 Db 956 DAUSRPKRELILFSGMARDPORLYVIQDERHLPSPDTSFYRALMDEEDMEDVDDA 1015
 Qy 1021 EYLVPQQGFFCPDPAPFAGGAGMVHRRHSRSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080
 Db 1016 DEYLVPQQGFF-----NSPST-----SRTPLLSSLS 1041
 Qy 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLORYSEDPTVPLPSET--DGYVAPLTCSPQPE 1138
 Db 1042 ATSN---NSTVACINRNGSCRVKEDAFQRYSSDPTGAVTEDNIDDAFL-----EVPE 1091
 Qy 1139 VYNQPDVVRPQPPSPREGPLPAAPAGATLERAKTLPKGNVVKDVFAFGAVENPEYL- 1197
 Db 1092 VYNQ-SVPRKPAQSVQNVFVTHNQPLHP-----APGRDLHYQN---PHNAVGNPNYLN 1140
 Qy 1198 TPQGAAPQHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPSTFKGTP 1241
 Db 1141 TAO-----PTCLSSGFSNPAWIKGSHQMSLDNDPDYQDFFPKPKNGIFKG-P 1190
 Qy 1242 TAENPEYGLDVP 1254
 Db 1191 TAENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1995
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <BE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <BE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.4%; Score 3094.5; DB 1; Length 1223;
Best Local Similarity 48.7%; Pred. No. 5e-119;
Matches 632; Conservative 173; Mismatches 346; Indels 147; Gaps 27;

QY 8 RWGLLLALLPPGAA-----STVCTGTDMLRLPASPETHDMLRLHYGCGVQVQNL 61
DB 13 RGAALVLLLLGVALCSAVEBEKKVCGQTNNKLTLGHVEDHFTSLQRMVNNCEVLSNLE 72
QY 62 LTVLPTNASLSFLQDIOEVGYVLIHNOVQVPLQRLRVGCTQLFEDNYALVLDNGD 121
DB 73 ITVEHNRDITFLKTIQEVAGYVLIHNMVDVPLENLIQIRGNVLYDNSFALVLSNYH 132
QY 122 PLNNTTPTVGTASPGGLRELQSLRSLTEILKGGVLTQRPOLCYQDTILWKDIFHKNQAL 181
DB 133 -MNKTQ-----GLRELPMKRLSELNGGVKLSNNPKLCNMDTVLWMDIITSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKGSRGWSSBDCSLTRTVCAAGCA-RCKGLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNECTEDHCWGAGEQNCQTLTKVICAQQCSCGRGKVPSPDCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPGRYTCGSCVTAC 299
DB 243 CAAGCTGPRSDCLACKKFRDDATCKDTCPLVLYNPTTYMDVNPBGYSFGATCVR 302
QY 300 PYNLSTDVGSCTLVCPPLHNEVTAEDGTQRCCKSPCARVCVGLGMQVIKANSKFIGI 359
DB 303 PHNVVYTDHSCVRSNCNTDIYEV-EENGVRKCKCDGLSKVCNGVIGIGELKGLS-INA 360
QY 360 TELE-PAGCKKIFGSLAFLESFGDPSANTAPLOPQLOVFFTEBITGYLYISAWPDS 418
DB 361 TNIDSFNCKTKINGDSILPVAFLDGAFKTLPLDPKKLDVFRVTKVKEISGFLLIQAWPDN 420
QY 419 LPDLSVFQNLQVIRGRILHNGVSLTLQGLIGISWGLRSRLRSLGSLALHHTHLCFVH 478
DB 421 ATDLYAFENLEIRGTRKQGVSLAVVNLKIOSLGLRSLKETSDGDIAMKKNLCYAD 480
QY 479 TVPWQDLFRNPQOALLHTANRPECEVGEGLACHOLCARGHCWPGPTQCVCNCSQFLRQ 538
DB 481 TMNWSLSFATQSQTKLIQNNKNDCTADRHVCDPLCSDVCGCGPGPFCHFCRFFSRQK 540
QY 539 ECVEECRVLQGLPREYVVARHCLFCHPECPQNG---SVTCFGPEADQCACAHYKDPFP 595

Db 541 ECVKQCNILQGBREFEERDSKCLPCHSECLVQNSTAYNTTSCSGPGDPDKCAHFIDGPH 600
QY 596 CVARCPGKVPDLISYMPKFPDEEGACOPCPNCTHSCVDLDDKGCPCAPQASPLTSIV 655
Db 601 CVKACFAGVLGENDTL-VWKYADANAVCQCHPNCTRGCKGPGLEGCP---NGSKTSPSIA 656
QY 656 SAVV-GILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRI 714
Db 657 AGVVGGLLCLVVVGLGIGLYLRR-HIVKRTLLRLLQERELVEPLTPSGEAPNQHURI 715
QY 715 LKTELKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 774
Db 716 LKETEFKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVM 775
QY 775 AGVSGPVSRLLGICLTSTVOLVQTMPLYGCLLDHYRENNRGRIGSDLLANWCQIAKMS 834
Db 776 ASVDNPHVCCLLIGICLTSTVOLITQTMPLYGCLLDYIREHKDNIGSQVLLNWCQIAKMN 835
QY 835 YLEDVRLVHRDLAARNVLRKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESI 894
Db 836 YLEERRLVHRDLAARNVLRKSPNHVKITDFGLAKLLGADSEKEYHAEGKVPKIKWMALESI 895
QY 895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMI 954
Db 896 LHRIFTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPPICTIDVYMI 955
QY 955 MVKCMWIDSECRFRFRELSESRMARDPQRFVVIQ-NEDLGPASPLDSFYBSLLEDDDD 1013
Db 956 MVKCMWIDADSKPRFRELTAEFSEKMDPPRYLVIOGDERMHLPSPTDSKFYTLMBEED 1015
QY 1014 MGDVDAEYLVPOQGFPCDPAPAGAGVMVHRHRSRSTRSGGDLTLGLEPSEEEAPRS 1073
Db 1016 MEDIVDAEYLVPHQGF-----NSPST-----SRT 1041
QY 1074 PL-----APSEAGAGSVDFDGLGMAAKGLQSLPTHDPPLORYSBDPTVPLPSET--DG 1126
Db 1042 PLLSSLSATSNNATNCID-----RNGQGHVREDSPVQRYSSDPTGNFLEESIDDG 1093
QY 1127 VYAPLTCSPQEVVNVQVDRPQPPSPREGPLPAARAGATLERAKTLSPKNGVGVKDFV- 1185
Db 1094 FL-----PAPEYVNO--LMPKKPS-----TAMVQNOIYNNISL 1124
QY 1186 -----AFGGAIVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQ----- 1225
Db 1125 TATSKLPMDSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVPSESPYWIQSGNHQI 1176
QY 1226 --DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 1177 NLNDNPYQODFLPNETKPNGLLKVPAAENPEYLRVAAP 1214
RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N.
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
C:Note: sequence extracted from NCBI backbone (NCBI.P126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match	43.5%	Score	2966.5	DB 2	Length	1308								
Best Local Similarity	45.2%	Pred.	No. 8.7e-114											
Matches	609	Conservative	184	Mismatches	380	Indels	173	Gaps	28					
Qy	9	WGLLLALLPGAA-----STOVC	GTDMKRLR	PASPETH	LDMLRHL	YOGCQV	OGNLELTY	64						
Db	8	WVWSLLVAAGT	VQPSQS	VCAGTEN	KSLS	DLQOYR	ALRKYKYE	CEVVMGNLS	67					
Qy	65	LPTNASLSFLQD	IOEQVGY	VLIAHQV	QVPLQ	RLRIVR	GTQLP	FEDNYAL	VDLNGDPLN	124				
Db	68	IEHRDLSFLRS	VREVTGY	VLVALNQ	RFYPL	PLENLR	IRGT	KLVEDYAL	ALFLNYRKDG	127				
Qy	125	NTTFVTGAS	PGGLRE	LQRLS	TEILK	GGVLIR	QNPQL	CYQDTIL	LWKDIF	FHKNQOLA	184			
Db	128	NF-----	GLQEL	KLNLTE	ILNGVY	VDQNK	FLCYAD	TIHQD	IVRN	WPSNLT	178			
Qy	185	DTNRSRACH	PCSPMCK	SRCWG	SESDC	OSLTR	TRTV	CAGGC	-ARCK	GPLP	DTDCHE	QCAAG	243	
Db	179	STNGSSCG	GRCHK	SGCTG	-RCW	GPTE	NHCOT	LRTR	VCAEQ	DGRCY	GPVSD	CCHE	CAGG	237
Qy	244	CTGPKG	SDCLAC	LHFNH	SGICEL	HCPAL	VNTDT	PESM	NPGE	YRTF	FGAS	CVTAC	PVNY	303
Db	238	CSGPK	DTDC	FACM	NFDS	GACV	TQCP	QTFV	YNPT	FQLE	HN	FN	AKYTY	297
Qy	304	LSTDV	SGCTLL	VCPL	HNQ	EVTA	EDGT	RCER	KSKP	CA	RVCY	GLG	MOYI	363
Db	298	V-VDS	SCVRAC	PSK	MEV-E	ENG	IKMCK	PCTD	IC	PKAC	DIG	GTG	SLMS	355
Qy	364	PAGCK	KIFG	SLA	FLP	ES	FDGP	PAS	NATAP	LQ	PE	LOV	FE	423
Db	356	F	INCTK	ING	NLFI	LVTG	I	GHG	DPYNA	I	EAID	PEK	LVN	415
Qy	424	VFQNL	QVIR	GRIL	HN	GA	YSLT	LOG	LGI	SW	LG	RLS	REL	483
Db	416	VFS	MLVTI	GER	VL	SG	LSLL	ILK	QO	GIS	T	S	FQ	475
Qy	484	QLFRNP	Q	ALLHT	ANR	PE	DCV	CG	EG	LACH	QL	CAR	GH	543
Db	476	TLF	STIN	QIR	VIR	DR	NR	KA	EN	CTA	EG	MV	CN	535
Qy	544	CRVL	Q	GLP	PRE	VN	AR	HCL	P	CH	PE	QCP	-QNG	602
Db	536	CNLY	DG	SF	REF	ENG	SI	CV	ED	PQ	CE	KEM	DGL	595
Qy	603	GVP	ED	LS	MP	IT	KFP	DE	EG	A	CQ	PC	IN	651
Db	596	GLQ	ANS	F--	IF	K	YAD	P	RE	CH	P	NC	TQ	652
Qy	652	TSI	VS	AVV	-GILL	VV	LV	GW	FG	IL	I	KR	QO	710
Db	653	--	IAAG	VIG	L	FIL	V	IGL	F	AV	V	RR	KI	708
Qy	711	QMR	IL	K	ET	EL	R	KV	K	VL	G	S	AG	770
Db	709	QLR	IL	K	ET	EL	R	KV	K	VL	G	S	AG	768
Qy	771	AYM	AG	V	G	S	P	V	S	L	L	I	G	830
Db	769	ALIN	AS	MD	H	PH	L	V	LL	G	V	S	P	828
Qy	831	KGMS	Y	ED	V	R	L	V	H	R	D	L	A	890
Db	829	KGM	Y	LE	R	L	V	H	R	D	L	A	R	888
Qy	891	LES	I	L	R	R	F	T	H	S	D	V	S	950
Db	889	LEC	I	H	R	K	F	T	H	S	D	V	S	948
Qy	'951	VYMI	T	W	K	C	M	I	D	S	R	P	K	1009
Db	949	VYMW	V	K	C	M	I	D	S	R	P	K	E	1008
Qy	1010	EDDD	M	G	D	L	V	D	A	E	Y	L	V	1069

[illegible]

Db 123 YQK-NPSSP--DYVQVGLKQLQLSNLTLLSGGVKVSHPNPLLNVTETINWWDIVDKTSNP 179
QY 180 ALTLIDTNRACHPCSPMKGSRGWESSEDCQSLTRTVCAGC-ARCKGPLPTDCCHE 238
Db 180 TMLNLI PHAFERQCKDHGCVNGSCWAPGPGCKQKFKLLCAEQCNRRCRGPKEPDCNE 239
QY 239 QCAAGCTGPHXSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
Db 240 HCAGGCTGPRATDCLACRDNDDGCTCTPPPKXIYDIVSHQVVDNPNIKYTFGAACVKE 299
QY 299 CPVNYLSTDVSGCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQVYKANSKFIG 358
Db 300 CPSNYVTE-GACVRSACGMLVD-ENGRRCKPCDGVCPKCDGIGIGSL-SNTIAVN 356
QY 359 ITEL-EPAGCKTIFGSLAFIPESFDGDPASNTAPLOPEQLQVFETLEETGYLISAWPD 417
Db 357 STNIRSPNCTKINGDIILNRNPEGDPHYKIGTMDPEHLNLTVEITGYLIVMWPE 416
QY 418 SLPLSVFQNLQVIRGILHNGAYS-LTLQGLGISWGLSLRSLGSLALIHNTLCP 476
Db 417 NMTSLSVFQNLQVIRGILHNGAYS-LTLQGLGISWGLSLRSLGSLALIHNTLCP 476
QY 477 VHTVPDQLFRNPHOALLHTANRPEDECVGEGACHOLCARGHCGPPTQCVNCSQFLR 536
Db 477 ANTNWRRLFRSEDSQSYDART-----ENQTCNNECEDSGCW-PPTWCWSCULHVR 528
QY 537 GQECVEBCEVRLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFC 596
Db 529 GGRVASCNLLQGEPREAQDGRVCVQCHQELVQTDLSLTCVGPANCKSAHFQDGPQC 588
QY 597 VARPSPGVKPLSYMPTWKPEDEGACQPCPINCTHSCVDLDDKGCAPABORAPLTSIVS 656
Db 589 IPRCPHGILGDDTL-IWKYADKMGQCPQCHQNTQCGSPGLSGCRGD-IVSHSLAVG 646
QY 657 AVVGILLVVLGVVFGILIKRRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQWRLK 716
Db 647 LVSGLLITVIALIVVLLRRRIK-RRIRICLLQKELVEPLTPSGAMPNQAQWRLK 705
QY 717 ETELKVKVLGSGAGFTYKIGIPIPDGENVKIPVAIKVLRNTPSPKANKELIDEAAYWAG 776
Db 706 ETEFKDRVLGSGAGFTYKIGIPIPDGENVKIPVAIKVLRNTPSPKANKELIDEAAYWAG 765
QY 777 VGSFYVRLIGICLTSTVQLVTOLMPYGCCLLDHVRENRLGSGODLLNWCQIAKMSYL 836
Db 766 VDPHVCRLGICLTSAVQLVTOLMPYGCCLLDVYRQHERICGQWLLNWCQIAKMSYL 825
QY 837 EDVRLVHRDLAARNVLKSNHYKITDFGLARLLIDETEHADGGKVPKMWALESLR 896
Db 826 EERHLVHRDLAARNVLKSNHYKITDFGLARLLIDETEHADGGKVPKMWALESLR 885
QY 897 RRETHQSDVMSYGVVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMV 956
Db 886 WTYTHQSDVMSYGVVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMV 945
QY 957 KCMWIDSECPRELVSFESRWARDQRFVITQNEDLGPASLDSTFYRSLLDDMDG 1016
Db 946 KCMWIDSECPRELVSFESRWARDQRFVITQNEDLGPASLDSTFYRSLLDDMDG 1000
QY 1017 LVDAEYLVPOQGFCCPDPPAGGAVHVRHRSSTSGGDLTLGLEPSEEAAPRPLA 1076
Db 1001 VDAEYLVPOQGFCCPDPPAGGAVHVRHRSSTSGGDLTLGLEPSEEAAPRPLA 1023
QY 1077 PSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTV-PLPSETDGYVAPLTCSP 1135
Db 1024 PPTGH-----PVRENSITLRLNISDTQNALEKDLGH----- 1055
QY 1136 QPEYVNPQVPRQP-----PSPRE-----GLP-AARPAGATLERAKTLSPGKNGWK 1182
Db 1056 --EYVNPQGETSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSRPEYLNQNSL-- 1111
QY 1183 DVAFAGAVENPEVLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTP 1242
Db 1112 -PLVSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPA 1148

QY 1243 AENPEYLIG 1250
Db 1149 AENLEYLG 1156

RESULT 9

A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C;Accession: A36223; I59164
R;Kraus, M.H.; Iseling, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A;Title: Isolation and characterization of ERB3, a third member of the ERBB/epidermal
A;Reference number: A36223; MUID:90083234; PMID:2687875
A;Accession: A36223
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1342 <KRA>
A;Cross-references: GB:M29366
R;Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A;Title: Molecular cloning and expression of another epidermal growth factor receptor-
A;Reference number: I59164; MUID:90311312; PMID:2164210
A;Accession: I59164
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A;Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841
C;Genetics:
A;Gene: GDB:ERBB3; HER3
A;Cross-references: GDB:119880; OMIM:190151
A;Map position: 12q13-12q13
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol-
C;Keywords: ATP; phosphotransferase
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif

Query Match 35.4%; Score 2414.5; DB 2; Length 1342;
Best Local Similarity 40.4%; Pred. No. 2.7e-91;
Matches 531; Conservative 195; Mismatches 453; Indels 135; Gaps 34;

QY 10 GLLALLPPGAA--STQVCTGDMKRLPASPETHLDMLRLHYQGCVVQGNLELYLPT 67
Db 11 GLLFSLARGSEVNGSQAQVCPGTLNGLSVTGDENQYQTLVKLYERCEVMGNLEIVLTGH 70
QY 68 NASLSFLDIQIEVOGVVLIHQNQVROVPLQRLRIVRGQTQLFEDNYALAVLDNGDPLNNTT 127
Db 71 NADLSFLQVIREVTGVVLMVAMNEFSLPLNLRVVRGTQVYDGKFAIFVM-----LNYNT 125
QY 128 PVTGASPGGLRELQLRSLTEILKGVLIQNPOLCQVDITLWKDIHKNQLALTLDITN 187
Db 126 ----NSHALRQLRLTQLTEILSGGVYIEKNOKLCHMDTIDWRDIVDRD---AEIVKVD 178
QY 188 RSRACHPCSPMKGSRGWESSEDCQSLTRTVCAGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCFPCEVCKG-KCWGPGSEDCQTLTKTICAPQCNHCGFCGPNPQCCHDCGAGCSG 237
QY 247 PKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLSLT 306
Db 238 PDQTDCAFCHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLSLT 296
QY 307 DVGSCCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQVYKANSKF--IGITELE- 363
Db 297 DQTSVCRVACPDKNEVD-KNGLKMECPGGLCPKACEGTG-----SGSRFQTVDDSNIDG 350
QY 364 FAGCKIFGSLAFPLPESFDGDPASNTAPLQPBQLQVFETLEETGYLISAWPDSLPDLS 423
Db 351 FVNTKILGNLDFLTGLNGDPMHKIPALDPEKLVNFRVREITGYLNTQSPPHMHNFS 410
QY 424 VFQNLQVIRGILHNGAYS-LTLQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTVPW 482

Db 759 LAVGSLDHAHIVRLGLCPGSSLOLVTOYLPLGSLLDHVKHRETLPQQLLNWGVQIAK 818
Qy 832 GMSYLEDLVRLVHRLDAAARNVLKSPNHYKIDFGLARLLDIDETEYHADGGKVPKWMAL 891
Db 819 GMYLEEHSMVHRDLALRNWMLKSPSQVQVADFGVADLLPDDKQLLHSEAKTPIKWMAL 878
Qy 892 ESILRRPFTHOSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPPOPPICHTDV 951
Db 879 ESHFGKYTHOSDVMSYGVTVWELMTGAEYAGLRLAEIPDLLEKGERLAQPOICHTDV 938
Qy 952 YIMVVKWIMIDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLPASPLDSTFVRSLLD 1011
Db 939 YMWVVKWIMIDENIRPTFEKELANFTMRDPPRYLVIKRAS-GFGTP--PAAPSPVIT 995
Qy 1012 DDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLSESEF--- 1068
Db 996 KEL-----QEALEPEL-----DLDLDLEAESEGLA 1021
Qy 1069 -----EAPRSLAPSEG-----AGDVFDGDLGMGAAGLQSLPT 1103
Db 1022 TSLGALSPLFTGLTRPRGSQLSPSSGYMPNMNQSSLGAEACLSAVLGGREQFSRDISL 1081
Qy 1104 HDPSPLQKYSDDPTVLPSETDGVV---APL-----TC-----SPOPE---VYNOP 1143
Db 1082 H-PIPRGR-----PASESEGHVTCSEAELOKVKVCSRSRSRSPRFGDSAYHSQR 1133
Qy 1144 DVRRPPSPREGP-----LPAAPAGATLERAKTLSP-GKNGV-----KDVF 1185
Db 1134 HSLLTPTVPLSPGLEEDGNGVYMPDTHLRGSSSREGTLSSVGLSSVLGTEDEED-- 1191
Qy 1186 AFGAVENPEYLTPOGGAAPQHP 1210
Db 1192 -----EYEMNKRKRGSP-PRPP 1209

RESULT 11
TVFLV
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro-
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NLI>
A;Cross-references: GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ENV>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.28; Pred. No. 3.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCACAHYKDPDFCVARCPGKVPDLSYMPKFPDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCMCAHFIDGHCVCACGAGVLGENDTL-VMKYADANAVCOLCHPNCTRCKGP 116
Qy 638 DDKGCPAEQASPLTSSIVAV-GILLVVVLGVVFGILIKRQOKRKTYMRLLOTEL 696

Db 117 GLEGP---NGSKTPSIAAQVVGGLCLVVGIGIGLYLRRR-HIVKRKTLRRLQREL 172
Qy 697 VEPLTPSGAMPNQAOMRILKETELRKVKVLGSGAFGVYKGIWIPGCVKIPIVAIKVL 756
Db 173 VEPLTPSGEAPNAHLRLKETEFKVKVLGSGAFGVYKGLWIPGEKVKIPIVAIKEL 232
Qy 757 ENTSPKANKEILDEAYVMAGVSGPYVSRLLGICLTSTVQLVTOLMPYVGCLLDHVRENRR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQIMPYVGCLLDVIREHKN 292
Qy 817 LGSODLLNMCWOLAKGMSVLEDLVRLHDLAARNVLKSPNHYKIDFGLARLLDIDETE 876
Db 293 IGSQYLLNMCVQIAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLAKLGLGADEKE 352
Qy 877 YHADGGKVPKIMWALSIILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 353 YHAEGGKVPKIMWALSIILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 412
Qy 937 KGERLPQPPICITDVTMIMVKWIMIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPQPPICITDVTMIMVKWIMIDADSRPKFRELIAEFSKWARDPPRYLVITQGERMH 472
Qy 996 PASPLDSTFVRSLLDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSG 1055
Db 473 LPSTDSKFTVRLMEEDMEDIVDAEYLVPHQGF-----NSPST--- 513
Qy 1056 GGLTLGLPSEBEAPRSL-----APSEGAGSDVFDGLGMGAAGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSATSNNSATNCID-----RNGQHPVRESFVQ 550
Qy 1111 RYSEDPVLPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLE 1168
Db 551 RYSSDPTGNFLESIIDGFL-----PAPEYVNVQ--LMPKKPS----- 585
Qy 1169 RAKTLPSPKNGVVKDVF-----AFGAVENPEYLTPOGGAAPQHPPPAP 1213
Db 586 ----TAMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLA 633
Qy 1214 SPAFDNLVYWDQ-----DPE-----RGAPPSTFGKTPTAENPEYVGLDVP 1254
Db 634 KTVFESSPWIQSGNHQINLDPDYQODFLPNETKNGLLKVPAAENPEYLVRAAP 689

RESULT 12
TVVUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene fami
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <WAG>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin,
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28, 'W', 30-139, 'P', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

QY	720	LRKVKYLGSAFGTVYKGIWIPDGENKVPVIAIKVLRENTSPKANKEILDEAYVMAGVGS	779
Db	810	LRKGGVGLMGAFGRVYKGVWVPEGENKVPVIAIKELLKGTGAESSEFUREAYIMASEBH	869
QY	780	PYVSRLLGICLSTWOLVTQIMPYGCLLDHVRENCRGLGSDLLANWCIOAKGMSVLEDV	839
Db	870	VNLLKLLAVCMSQMMMLITQIMPLGCLLDYVRNRDRKIGSKALLNNWSTQIAKMSYLEBK	929
QY	840	RLVHRDLAARNVLVK--SPNHVKITDFGLARLLDIDETEHYHADGGKVPKIKWMALESILR	896
Db	930	RLVHRDLAARNVLRLLAGEDH---DFGLAKLLSSDSNEYKAAAGCKMPIKWLALCEICRN	985
QY	897	RRFTHGSDVMSGYVTWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMWV	956
Db	986	RVFTSKSDVWAFGVTIWEILLTFQORPHENIPAKDIPDLIEVGLKRPQBCISLDIYCTILL	1045
QY	957	KCHWIDSECPREPLVSEFSRMARQPVVITQINEDLG--PASPLDSTFYRSLLEDD--	1012
Db	1046	SCWHLDAARPTFKQLTVPFAEFARDPGRYLAIGDKFTRLPA-----YTSQDEKDLI	1098
QY	1013	-DMGDLVDAEYLVPOQGFPCPDPAFGAGMVHHRHSSTRSGGGDLTLGLEPSEEEAP	1071
Db	1099	RKLAPTTDGEATAKPPDYLOPKAALGPS-----HRTDCT-----DEMP	1137
QY	1072	-----RSLPAPSEAGSDVFDG---DLNGWAAKGLQSLPTHDPSPLOYSEDPVPLPS	1122
Db	1138	KLNRKYCKDPNKNKSGTDDERDSAREVGVGNLR-----LDLPV	1176
QY	1123	ETDGYVAPLTCSPQPEVVNQPDVPPSPREGPLPAARPAGATLERAKTLSPGKNGVVK	1182
Db	1177	DEDDYLMP-TCQPGPNNNNNM-----NPNQNNMAAVGVAAGYM-----	1214
QY	1183	DVFAFGAVENPEYL-----TPOGGAAPQH-----PPPAFSP-AFON	1219
Db	1215	DLIGVPVSVDNPEYLLNAQTLGVGESPIFTQTIGIPWGGPGTWEVKVMPGSGEPTSSDH	1274
QY	1220	LYYWD	1224
Db	1275	EYND	1279

Search completed: July 22, 2003, 09:09:23
Job time : 30.9062 secs